

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Computer Ltd.

OM protein - protein search, using sw model

Run on: December 13, 2002, 15:52:26 : Search time 35 seconds  
(without alignments)  
2566.758 Million cell updates/sec

Title: US-09-866-379a-10  
Perfect score: 2279  
Sequence: 1 MKAILTFLLTLTQSAFAOSEFELKLESVV1552HCVRAPTKATQLEMDVTPAMP 60

Scoring table: BLUSUMB2

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, z00047115 residues 671580

Total number of hits satisfying chosen parameters.  
Minimum DB seq length: 0  
Maximum DB seq length: 200000000  
Post-processing: Minimum Match 6%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL21.\*  
1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mhc.\*  
8: sp\_orqanelle.\*  
9: sp\_plaque.\*  
10: sp\_plant.\*  
11: sp\_rodent.\*  
12: sp\_virus.\*  
13: sp\_vertebrate.\*  
14: sp\_unclassified.\*  
15: sp\_virus.\*  
16: sp\_bacteriaph.\*  
17: sp\_archaea.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	DB ID	Description
1	2176	95.5	432	2 Q8RFF3	Q8RFF3 escherichia
2	2176	95.5	432	2 Q8RKH8	Q8RKH8 escherichia
3	2176	95.5	432	2 Q8RKH7	Q8RKH7 escherichia
4	2176	95.5	432	2 Q8RKH6	Q8RKH6 escherichia
5	2175	95.4	432	2 Q8RKH5	Q8RKH5 escherichia
6	2172	95.3	432	2 Q8RKH4	Q8RKH4 escherichia
7	2144	94.1	434	16 Q8XC29	Q8XC29 escherichia
8	959.5	42.1	441	16 Q8RFF6	Q8RFF6 yersinia pe
9	560	24.6	413	16 Q8RKH7	Q8RKH7 escherichia
10	547	24.0	413	16 Q8RKH6	Q8RKH6 escherichia
11	536	23.5	414	16 Q8RKH5	Q8RKH5 escherichia
12	227.5	13.3	419	2 Q8RKH4	Q8RKH4 escherichia
13	231	8.8	125	4 Q8RKH3	Q8RKH3 escherichia
14	146	6.4	531	5 Q8RKH2	Q8RKH2 escherichia
15	140.5	6.2	333	4 Q8RKH1	Q8RKH1 escherichia
16	137	6.0	416	5 Q19390	Q19390 caenorhabdi

17	135	5.9	423	4 Q9BTU7	Q9BTU7 homo sapien
18	131	5.7	542	5 Q9ULA2	Q9ULA2 leishmania
19	127	5.6	395	5 Q9V6B8	Q9V6B8 drosophila
20	126	5.5	432	3 Q8WZJ5	Q8WZJ5 aspergillus
21	126	5.5	465	3 Q00092	Q00092 aspergillus
22	126	5.5	684	5 Q00838	Q00838 leishmania
23	126	5.5	707	5 Q00839	Q00839 leishmania
24	124.5	5.5	447	5 Q9U5U3	Q9U5U3 drosophila
25	123.5	5.4	392	5 Q9V5V4	Q9V5V4 drosophila
26	123.5	5.4	447	5 Q9U5V1	Q9U5V1 drosophila
27	120.5	5.3	447	5 Q9U5U4	Q9U5U4 drosophila
28	120.5	5.3	447	5 Q9U5U5	Q9U5U5 drosophila
29	120.5	5.3	447	5 Q9U5U6	Q9U5U6 drosophila
30	120.5	5.3	447	5 Q9U5U7	Q9U5U7 drosophila
31	120.5	5.3	447	5 Q9U5U8	Q9U5U8 drosophila
32	120.5	5.3	447	5 Q9U5U9	Q9U5U9 drosophila
33	120.5	5.3	447	5 Q9U5U0	Q9U5U0 drosophila
34	120.5	5.3	447	5 Q9U5U1	Q9U5U1 drosophila
35	120.5	5.3	447	5 Q9U5U2	Q9U5U2 drosophila
36	120.5	5.3	447	5 Q9U5U3	Q9U5U3 drosophila
37	120.5	5.3	447	5 Q9U5U4	Q9U5U4 drosophila
38	120.5	5.3	447	5 Q9U5U5	Q9U5U5 drosophila
39	120.5	5.3	447	5 Q9U5U6	Q9U5U6 drosophila
40	120.5	5.3	447	5 Q9U5U7	Q9U5U7 drosophila
41	120.5	5.3	447	5 Q9U5U8	Q9U5U8 drosophila
42	120.5	5.3	447	5 Q9U5U9	Q9U5U9 drosophila
43	120.5	5.3	447	5 Q9U5U0	Q9U5U0 drosophila
44	120.5	5.3	447	5 Q9U5U1	Q9U5U1 drosophila
45	119.5	5.2	447	5 Q9U5U2	Q9U5U2 drosophila

ALIGNMENTS

RESULT 1

Q8RKE0 Q8RKE0 PRELIMINARY: PRT: 432 AA.  
AC Q8RKE0:  
DT 01-JUN-2002 (TRMBLrel. 21, Created)  
DT 01-JUN-2002 (TRMBLrel. 21, Last sequence update)  
DT 01-JUN-2002 (TRMBLrel. 21, Last annotation update)  
DE periplasmic phosphohydrolase phosphohydrolase.  
GN APPA.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90368616; PubMed=2168385;  
RA Dassa J., Marek C., Boquet P.-L.L.;  
RT "The complete nucleotide sequence of the Escherichia coli gene appA  
reveals significant homology between pH 2.5 acid phosphatase and  
glucosyl-phosphatase."  
RL J. Bacteriol. 172:5497-5500(1990).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94054596; PubMed 1429631;  
RA Ostanin K., Harms E.H., Stevis P.E., Kuciel K., Zhou M.M.,  
Van Etten R.L.;  
RT "Overexpression, site-directed mutagenesis, and mechanism of  
Escherichia coli acid phosphatase."  
RL J. Biol. Chem. 267:22839-22846(1992).  
DR EMBL: I03370; AAA00002.1;  
KW Hydrolyase  
SQ SEQUENCE 432 AA, 4697 MW, 580563.54692EABF CRC64;  
Query Match 95.5%, Score 2176; BB 2; Length 432;  
Post-local Similarity 97.9%; Pred. No. 3.4e-170;  
Matches 423; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 MKAILTFLLTLTQSAFAOSEFELKLESVV1552HCVRAPTKATQLEMDVTPAMP 60  
|||||

```

Db      1  MKAILLPFLSLIPITPQSAFAQSEPEIKLESVVIVSARHVRAPIKATQIMQDVTTPAMP 60
QY      61  TWPVKAGELTPRGGELIAYLGHYWRQRIVADCLLPKGRCPQSGQVALLIAWDERTKIGE 120
Db      61  TWPVKAGELTPRGGELIAYLGHYWRQRIVADCLLPKGRCPQSGQVALLIAWDERTKIGE 120
QY      121  AFAAGLAHPCALIVHTQADISSPPLENDUKTGVGQILINANVTUATLIRAGGSIAIDFTGH 180
Db      121  AFAAGLAHPCALIVHTQADISSPPLENDUKTGVGQILINANVTUATLIRAGGSIAIDFTGH 180
QY      181  YQAFRELERVNFPPQSNLZTEREKQDSSTSLGALSFSEKVSALVSLIYAVSLASMLT 240
Db      181  YQAFRELERVNFPPQSNLZTEREKQDSSTSLGALSFSEKVSALVSLIYAVSLASMLT 240
QY      241  EFTLQAAGMPGPGWGRITTSHOWNTLILSLNAGFELLOKTPVAPSPATPLLDLTKTA 300
Db      241  EFTLQAAGMPGPGWGRITTSHOWNTLILSLNAGFELLOKTPVAPSPATPLLDLTKTA 300
QY      301  LIPHPQKQAGVTLPTSVLEIACHDINLANIGALFNWTLPGQDNTPGSGELVFERW 360
Db      301  LIPHPQKQAGVTLPTSVLEIACHDINLANIGALFNWTLPGQDNTPGSGELVFERW 360
QY      361  KRLSNSQWIGVSVFQTLQOMRKTPLSLNTDPSVFKTLTAGFPFNAQGWSTASPTQ 420
Db      361  KRLSNSQWIGVSVFQTLQOMRKTPLSLNTDPSVFKTLTAGFPFNAQGWSTASPTQ 420
QY      421  LVNEARIPACSL 432
Db      421  LVNEARIPACSL 432

```

## RESULT 2

```

QRRK08
ID  QRRK08  PRELIMINARY:  PRF:  432 AA.
AC  QRRK08
DT  01-JUN-2002 (TEMBLRel. 21, Created)
DE  01-JUN-2002 (TEMBLRel. 21, last sequence update)
DE  01-JUN-2002 (TEMBLRel. 21, last annotation update)
DE  Periplasmic phosphoanhydride phosphohydrolase.
GN  APPA.
OS  Escherichia coli.
OC  Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC  Escherichia.
OX  NCBI_TaxID 562;
RN  [1]
RX  MEDLINE 90368616, PubMed 2168485;
RA  Dassa J., Marek C., Boquet P.-L.L.;
RT  "The complete nucleotide sequence of the Escherichia coli gene appA
RT  reveals significant homology between pH 2.5 acid phosphatase and
RT  glucose-1-phosphatase.";
RL  J. Bacteriol. 172:5497-5500(1990).
RN  [2]
RP  SEQUENCE FROM N.A.
RX  MEDLINE 93054596; PubMed 1429631;
RA  Oostanin K., Harms E.H., Stevis P.E., Kuo-Tei P., Zhou M.M.,
RA  Van Elteren R.;
RT  "Overexpression, site-directed mutagenesis, and mechanism of
RT  Escherichia coli acid phosphatase.";
PL  J. Biol. Chem. 267:22830-22836(1992).
DR  EMBL: L03373 AAA00005.1;
KW  Hydrolyase.
SQ  HYDROLASE. 432 AA; 46971 MW; 98553057800045 (CR04);

```

Query Match 95.5%; Score 2176; DB 2; Length 432;

Best local Similarity 97.9%; Pred. No. 34e 170;

Matches 423; conservative 0; Mismatches 9; Indels 0; Gaps 0;

```

QY      1  MKAILLPFLSLIPITPQSAFAQSEPEIKLESVVIVSARHVRAPIKATQIMQDVTTPAMP 60
Db      1  MKAILLPFLSLIPITPQSAFAQSEPEIKLESVVIVSARHVRAPIKATQIMQDVTTPAMP 60
QY      61  TWPVKAGELTPRGGELIAYLGHYWRQRIVADCLLPKGRCPQSGQVALLIAWDERTKIGE 120

```

```

Db      61  TWPVKAGELTPRGGELIAYLGHYWRQRIVADCLLPKGRCPQSGQVALLIAWDERTKIGE 120
QY      121  AFAAGLAHPCALIVHTQADISSPPLENDUKTGVGQILINANVTUATLIRAGGSIAIDFTGH 180
Db      121  AFAAGLAHPCALIVHTQADISSPPLENDUKTGVGQILINANVTUATLIRAGGSIAIDFTGH 180
QY      181  YQAFRELERVNFPPQSNLZTEREKQDSSTSLGALSFSEKVSALVSLIYAVSLASMLT 240
Db      181  YQAFRELERVNFPPQSNLZTEREKQDSSTSLGALSFSEKVSALVSLIYAVSLASMLT 240
QY      241  EFTLQAAGMPGPGWGRITTSHOWNTLILSLNAGFELLOKTPVAPSPATPLLDLTKTA 300
Db      241  EFTLQAAGMPGPGWGRITTSHOWNTLILSLNAGFELLOKTPVAPSPATPLLDLTKTA 300
QY      301  LIPHPQKQAGVTLPTSVLEIACHDINLANIGALFNWTLPGQDNTPGSGELVFERW 360
Db      301  LIPHPQKQAGVTLPTSVLEIACHDINLANIGALFNWTLPGQDNTPGSGELVFERW 360
QY      361  KRLSNSQWIGVSVFQTLQOMRKTPLSLNTDPSVFKTLTAGFPFNAQGWSTASPTQ 420
Db      361  KRLSNSQWIGVSVFQTLQOMRKTPLSLNTDPSVFKTLTAGFPFNAQGWSTASPTQ 420
QY      421  LVNEARIPACSL 432
Db      421  LVNEARIPACSL 432

```

## RESULT 3

```

QRRK07
ID  QRRK07  PRELIMINARY:  PRF:  432 AA.
AC  QRRK07
DT  01-JUN-2002 (TEMBLRel. 21, Created)
DE  01-JUN-2002 (TEMBLRel. 21, last sequence update)
DE  01-JUN-2002 (TEMBLRel. 21, last annotation update)
DE  Periplasmic phosphoanhydride phosphohydrolase.
GN  APPA.
OS  Escherichia coli.
OC  Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC  Escherichia.
OX  NCBI_TaxID 562;
RN  [1]
RX  SEQUENCE FROM N.A.
RX  MEDLINE 90368616; PubMed 2168485;
RA  Dassa J., Marek C., Boquet P.-L.L.;
RT  "The complete nucleotide sequence of the Escherichia coli gene appA
RT  reveals significant homology between pH 2.5 acid phosphatase and
RT  glucose-1-phosphatase.";
RL  J. Bacteriol. 172:5497-5500(1990).
RN  [2]
RP  SEQUENCE FROM N.A.
RX  MEDLINE 93054596; PubMed 1429631;
RA  Oostanin K., Harms E.H., Stevis P.E., Kuo-Tei P., Zhou M.M.,
RA  Van Elteren R.;
RT  "Overexpression, site-directed mutagenesis, and mechanism of
RT  Escherichia coli acid phosphatase.";
PL  J. Biol. Chem. 267:22830-22836(1992).
DR  EMBL: L03373 AAA00005.1;
KW  Hydrolyase.
SQ  HYDROLASE. 432 AA; 46971 MW; 750543401A9914A6 (CR04);

```

Query Match

Best local Similarity 97.9%; Pred. No. 34e 170;

Matches 423; conservative 0; Mismatches 9; Indels 0; Gaps 0;

```

QY      1  MKAILLPFLSLIPITPQSAFAQSEPEIKLESVVIVSARHVRAPIKATQIMQDVTTPAMP 60
Db      1  MKAILLPFLSLIPITPQSAFAQSEPEIKLESVVIVSARHVRAPIKATQIMQDVTTPAMP 60
QY      61  TWPVKAGELTPRGGELIAYLGHYWRQRIVADCLLPKGRCPQSGQVALLIAWDERTKIGE 120
Db      61  TWPVKAGELTPRGGELIAYLGHYWRQRIVADCLLPKGRCPQSGQVALLIAWDERTKIGE 120

```

QY 121 AFAAGLAPDCAITVHTQADTSPPDPLFNPIKTGVCOLDNANVTDAILEPAGSSTADFTGH 180  
 DB 121 AFAAGLAPDCAITVHTQADTSPPDPLFNPIKTGVCOLDNANVTDAILEPAGSSTADFTGH 180  
 QY 181 YGTAFRELEKVINFPQSNLCKLKKQDESSLTGALFSELKVSADNVSTGAVSLASMLT 240  
 DB 181 YGTAFRELEKVINFPQSNLCKLKKQDESSLTGALFSELKVSADNVSTGAVSLASMLT 240  
 QY 241 EIFLLQAAGMPPEGWGRITLSDHWNLLSLHINAAGLLAGTFFIVASERATLTLTKTA 300  
 DB 241 EIFLLQAAGMPPEGWGRITLSDHWNLLSLHINAAGLLAGTFFIVASERATLTLTKTA 300  
 QY 301 LIPHPKQKAYGVLTPTSVLFTAGHDNLNLANLGGALLELNTLTGQTFRTTTSSELVFERW 360  
 DB 301 LIPHPKQKAYGVLTPTSVLFTAGHDNLNLANLGGALLELNTLTGQTFRTTTSSELVFERW 360  
 QY 361 RLSDNSQWIVQSVLFTQLOQMDKTPLSLNTPPGQVKTITLACCERNAGGMSLAGFTQ 420  
 DB 361 RLSDNSQWIVQSVLFTQLOQMDKTPLSLNTPPGQVKTITLACCERNAGGMSLAGFTQ 420  
 QY 421 IVNEARIPACSL 432  
 DB 421 IVNEARIPACSL 432

RESULT 4  
 Q8RKD6 PRELIMINARY: PRT: 432 AA.  
 AC Q8RKD6  
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Periplasmic phosphoanhydride phosphohydrolase.  
 GN APPA.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OX NCBI\_TaxID:562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90368616; PubMed=2168385;  
 RA Basso J., Marck C., Boquet P.-L.L.;  
 RT "The complete nucleotide sequence of the Escherichia coli gene appA  
 RT reveals significant homology between pH 2.5 acid phosphatase and  
 RT glucose-1-phosphatase."  
 RL J. Bacteriol. 172:5497-5500(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94054596; PubMed 1429631;  
 RA Ostlund K., Hains E.H., Stevis E.E., Kacieli E., Zhou M.M.,  
 RA Van Etten R.L.;  
 RT "Overexpression, site-directed mutagenesis, and mechanism of  
 RT Escherichia coli acid phosphatase".  
 RL J. Biol. Chem. 267:22830-22836(1992).  
 DR EMBL: L03371; AAA00006.1; -.  
 KW Hydrolase.  
 SQ SEQUENCE 432 AA; 46971 MW; AF56041FAA13AA5 C664;

Query Match 95.4%; Score 2176; DB 2; Length 432;  
 Best local Similarity 97.9%; Pred. No. 340-170;  
 Matches 423; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAALIFELSLILPLTPQSAFAQSEPELKESSVIVVGHCVAPRTKATQIMQVTPDAMP 60  
 DB 1 MAALIFELSLILPLTPQSAFAQSEPELKESSVIVVGHCVAPRTKATQIMQVTPDAMP 60  
 QY 61 TWPKVGLKLTTPGSELATLGHYQKRLVALGALAKESFTSSQVALIAVDTERETKGE 120  
 DB 61 TWPKVGLKLTTPGSELATLGHYQKRLVALGALAKESFTSSQVALIAVDTERETKGE 120  
 QY 121 AFAAGLAPDCAITVHTQADTSPPDPLFNPIKTGVCOLDNANVTDAILEPAGSSTADFTGH 180  
 DB 121 AFAAGLAPDCAITVHTQADTSPPDPLFNPIKTGVCOLDNANVTDAILEPAGSSTADFTGH 180  
 QY 181 YGTAFRELEKVINFPQSNLCKLKKQDESSLTGALFSELKVSADNVSTGAVSLASMLT 240  
 DB 181 YGTAFRELEKVINFPQSNLCKLKKQDESSLTGALFSELKVSADNVSTGAVSLASMLT 240

QY 181 YGTAFRELEKVINFPQSNLCKLKKQDESSLTGALFSELKVSADNVSTGAVSLASMLT 240  
 DB 181 YGTAFRELEKVINFPQSNLCKLKKQDESSLTGALFSELKVSADNVSTGAVSLASMLT 240  
 QY 241 EIFLLQAAGMPPEGWGRITLSDHWNLLSLHINAAGLLAGTFFIVASERATLTLTKTA 300  
 DB 241 EIFLLQAAGMPPEGWGRITLSDHWNLLSLHINAAGLLAGTFFIVASERATLTLTKTA 300  
 QY 301 LIPHPKQKAYGVLTPTSVLFTAGHDNLNLANLGGALLELNTLTGQTFRTTTSSELVFERW 360  
 DB 301 LIPHPKQKAYGVLTPTSVLFTAGHDNLNLANLGGALLELNTLTGQTFRTTTSSELVFERW 360  
 QY 361 RLSDNSQWIVQSVLFTQLOQMDKTPLSLNTPPGQVKTITLACCERNAGGMSLAGFTQ 420  
 DB 361 RLSDNSQWIVQSVLFTQLOQMDKTPLSLNTPPGQVKTITLACCERNAGGMSLAGFTQ 420  
 QY 421 IVNEARIPACSL 432  
 DB 421 IVNEARIPACSL 432

RESULT 5  
 Q8RKD9 PRELIMINARY: PRT: 432 AA.  
 AC Q8RKD9  
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Periplasmic phosphoanhydride phosphohydrolase.  
 GN APPA.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OX NCBI\_TaxID:562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90368616; PubMed=2168385;  
 RA Basso J., Marck C., Boquet P.-L.L.;  
 RT "The complete nucleotide sequence of the Escherichia coli gene appA  
 RT reveals significant homology between pH 2.5 acid phosphatase and  
 RT glucose-1-phosphatase."  
 RL J. Bacteriol. 172:5497-5500(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94054596; PubMed 1429631;  
 RA Ostlund K., Hains E.H., Stevis E.E., Kacieli E., Zhou M.M.,  
 RA Van Etten R.L.;  
 RT "Overexpression, site-directed mutagenesis, and mechanism of  
 RT Escherichia coli acid phosphatase".  
 RL J. Biol. Chem. 267:22830-22836(1992).  
 DR EMBL: L03371; AAA00006.1; -.  
 KW Hydrolase.  
 SQ SEQUENCE 432 AA; 47033 MW; 0124893F73268175 C664;

Query Match 95.4%; Score 2175; DB 2; Length 432;  
 Best local Similarity 97.9%; Pred. No. 410-170;  
 Matches 423; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

```

Db 181 ROTAFRELERVINFPGSWCHIKRKKQDSQSLTQALPSELKVSADNVSLTQAVSLASMIT 240
QY 241 EFTLLQAQAQMPFPGWCHITISHQWNTLILSLINAQDFILLQRTPTVARSRAATPLLDLIKTA 300
Db 241 EFTLLQAQAQMPFPGWCHITISHQWNTLILSLINAQDFILLQRTPTVARSRAATPLLDLIKTA 300
QY 301 LTPHPQQAQYGVLTPTSVLFTAGHDTNLANIAGALELNWTLTGGDINTPTGGELVFERW 360
Db 301 LTPHPQQAQYGVLTPTSVLFTAGHDTNLANIAGALELNWTLTGGDINTPTGGELVFERW 360
QY 361 RRLSINSQWLOVSIVFQTLQOMRKPDLSTNTPPGVEVKLLAGSEERNAQCMCSLAGTQ 420
Db 361 RRLSINSQWLOVSIVFQTLQOMRKPDLSTNTPPGVEVKLLAGSEERNAQCMCSLAGTQ 420
QY 421 IVNEARIPACSL 432
Db 421 IVNEARIPACSL 432

RESULT 6
Q8XKJ5
ID Q8XKJ5 PRELIMINARY: PRT: 432 AA.
AC Q8XKJ5
DT 01-JUN-2002 (Tremblrel, 21, Created)
DT 01-JUN-2002 (Tremblrel, 21, Last sequence update)
DE periplasmic phosphoanhydride phosphatase.
GN APFA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
FN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE 90468616; PubMed=2168185;
RA Bessa J., Marrk C., Boquet P., Lili?
RT "The complete nucleotide sequence of the Escherichia coli gene apfa
RT reveals significant homology between pH 2.5 acid phosphatase and
RT glucose-1-phosphatase."
RI J. Bacteriol. 172:5497-5500(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE 94054596; PubMed=1429643;
RA Ostlin K., Bains E.H., Stevis P.E., Kuciel R., Zhou M.M.,
RA Van Etten R.L.?
RT "overexpression, site-directed mutagenesis, and mechanism of
RT Escherichia coli acid phosphatase."
RI J. Biol. Chem. 267:22840-22846(1992).
RW EMBL: L03375; AAA00007.1; ?
KW Hydrolyase.
SQ SEQUENCE 432 AA: 46990 MW: 951F304EA9A1AM7C GRC64;

Query Match
Best Local Similarity 97.98; Poses 2172; Db 2; Length 432;
Matches 424; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 MKAILLFLPSLLILPTPSAFAQSEPELKIESVIVSPHVPAPTKATLQGVTPDAMP 60
Db 1 MKAILLFLPSLLILPTPSAFAQSEPELKIESVIVSPHVPAPTKATLQGVTPDAMP 60
QY 61 TWPKRLDELTPRGCHTAYTCHYWRQRTVAGSHLPKAGTQSGGVAIALVIEPTKGE 120
Db 61 TWPKRLDELTPRGCHTAYTCHYWRQRTVAGSHLPKAGTQSGGVAIALVIEPTKGE 120
QY 121 AFAAGLAPVAVIIVTQADTSPPQVFNPIKTCVQVQVNNAVTDAIIPAGASSTADPTCH 180
Db 121 AFAAGLAPVAVIIVTQADTSPPQVFNPIKTCVQVQVNNAVTDAIIPAGASSTADPTCH 180
QY 181 YOTAFRELERVINFPGSWCHIKRKKQDSQSLTQALPSELKVSADNVSLTQAVSLASMIT 240
Db 181 YOTAFRELERVINFPGSWCHIKRKKQDSQSLTQALPSELKVSADNVSLTQAVSLASMIT 240
QY 241 EFTLLQAQAQMPFPGWCHITISHQWNTLILSLINAQDFILLQRTPTVARSRAATPLLDLIKTA 300

```

```

Db 241 EFTLLQAQAQMPFPGWCHITISHQWNTLILSLINAQDFILLQRTPTVARSRAATPLLDLIKTA 300
QY 301 LTPHPQQAQYGVLTPTSVLFTAGHDTNLANIAGALELNWTLTGGDINTPTGGELVFERW 360
Db 301 LTPHPQQAQYGVLTPTSVLFTAGHDTNLANIAGALELNWTLTGGDINTPTGGELVFERW 360
QY 361 RRLSINSQWLOVSIVFQTLQOMRKPDLSTNTPPGVEVKLLAGSEERNAQCMCSLAGTQ 420
Db 361 RRLSINSQWLOVSIVFQTLQOMRKPDLSTNTPPGVEVKLLAGSEERNAQCMCSLAGTQ 420
QY 421 IVNEARIPACSL 432
Db 421 IVNEARIPACSL 432

RESULT 7
Q8XC29
ID Q8XC29 PRELIMINARY: PRT: 444 AA.
AC Q8XC29
DT 01-MAR-2002 (Tremblrel, 20, Created)
DT 01-MAR-2002 (Tremblrel, 20, Last sequence update)
DT 01-JUN-2002 (Tremblrel, 21, Last annotation update)
DE Phosphoanhydride phosphatase, pH 2.5 acid phosphatase,
DE periplasmic.
GN APFA OR 21397 OR EGS1136.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=83334;
FN [1]
RP SEQUENCE FROM N.A.
RX STRAIN O157:H7 / EHL933 / ATCC 700927;
RX MEDLINE 21674936; PubMed=11206551;
RA Perna N.T., Plunkett G., III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Green J., Kirkpatrick H.A.,
RA Kosfai G., Haykett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch K.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7."
RI Nature 409:529-533(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN O157:H7 / KMD 9509562;
RX MEDLINE 21156241; PubMed=11250796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Take T.,
RA Iida T., Takami H., Honda T., Sasakiwa C., Ogasawara N., Yasumura T.,
RA Kubara S., Shiba T., Battori M., Shitayama H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12."
RI DNA Res. 8:311-322(2001).
RW EMBL: AF005292; AAC55528.1; ALL-INT1.
RW EMBL: AP002554; BAB4559.1; ?
RT Interpro: IPR000560; HSAc_phosphatase.
RX P433: PF03426; acid phosphat; 1.
RX PROSITE: PS00616; HIS-ACID-PHOSPHAT_1; 1.
RX PROSITE: PS04778; HIS-ACID-PHOSPHAT_2; 1.
KW Complete phosphatase.
SQ SEQUENCE 444 AA: 47437 MW: F192DF7B16A9F974 99064;

Query Match
Best Local Similarity 94.19; Poses 2144; Db 10; Length 444;
Matches 419; Conservative 0; Mismatches 14; Indels 2; Gaps 14;

QY 1 MKAILLFLPSLLILPTPSAFAQSEPELKIESVIVSPHVPAPTKATLQGVTPDAMP 60
Db 1 MKAILLFLPSLLILPTPSAFAQSEPELKIESVIVSPHVPAPTKATLQGVTPDAMP 60
QY 59 TWPKRLDELTPRGCHTAYTCHYWRQRTVAGSHLPKAGTQSGGVAIALVIEPTKGE 118
Db 59 TWPKRLDELTPRGCHTAYTCHYWRQRTVAGSHLPKAGTQSGGVAIALVIEPTKGE 118
QY 61 TWPKRLDELTPRGCHTAYTCHYWRQRTVAGSHLPKAGTQSGGVAIALVIEPTKGE 140
Db 61 TWPKRLDELTPRGCHTAYTCHYWRQRTVAGSHLPKAGTQSGGVAIALVIEPTKGE 140

```







```

247 QY QACMP--PPWCPPTDSSHWRTTISLHNAQTLQRTPEVARSATPLDLIKALPTH 404
      : : : : : : : : : : : : : : : : : : : : : : : : : :
232 DB WSNPLDLSLACKKITQACITALLPIFENYDINIVLVAQKASVFNAMIDVARKH 291
      : : : : : : : : : : : : : : : : : : : : : : : : : :

305 QY PPKQAYGVLTPTSVLFIAGHTNINLANGLALFINNTIDAPDNTPAQELVFVRWR 362
      : : : : : : : : : : : : : : : : : : : : : : : : : :
292 DB ANINPVW-----LLVADHTNIAMVPTIMNFSLDYSNGNIPGSSIVLEWRK 343
      : : : : : : : : : : : : : : : : : : : : : : : : : :

363 QY LSNSSOWLQVSLVEQI--LQGMPTAFETISLNTDPEVEL- LAGTEPNAQMG-- 413
      : : : : : : : : : : : : : : : : : : : : : : : : : :
344 DB RAENA--LQGSIFQAQIGIDUR-----SLATPDAHPEMLEGWRHAFGRTHVSLTAPF 495
      : : : : : : : : : : : : : : : : : : : : : : : : : :

414 QY --SLAGTQIVNEARIPAGSL 432
      : : : : : : : : : : : : : : : : : : : : : : : : : :
396 DB GAATTAIGORIDRSSAPAVAM 416
      : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT	3
Q9BZG2	
CDS	PREDIMINARY, FET, 426 AA.
AC	Q9BVG2;
DT	01-JUN-2001 (TrEMBL;rel 17, Created)
DI	01-JUN-2001 (TrEMBL;rel 17, Last sequence update)
PT	01-MAR-2002 (TrEMBL;rel 20, Last annotation update)
FT	Acid phosphatase.
GN	ACPT.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OX	Mammalia; Furberia; Primates; Catarrhini; Homnidae; Homo.
NCBI_TaxID=9606;	
[1]	
RN	
RP	SEQUENCE FROM N.A.
XP	MEDLINE=2130973; PubMed=11414767;
YF	Yusef G M., Diamandis M., Iung K., Diamandis E.P.;
FT	"Molecular cloning of a novel human acid phosphatase gene (ACPT) that
RT	is highly expressed in the testis.";
FL	Genomics 74:385-395(2001)
JR	EMBL; AF321918; AAK09393.1; -.
DR	HSPD_P15309; 2HPA.
IRP	InterPro: IPRO00560; HisAc_phosphatse.
PR	Fram; PF00328; acid_phosphat_1.
DP	PROSITE; PSM0116; HIS_ACID_PHOSPAT_1; 1.
DF	PROSITE; PSM00778; HIS_ACID_PHOSPAT_2; UNKNOWN_1.
SO	SEQUENCE 426 AA; 45089 MW; BE930398041IP061_CPE64;

QY	67	GLTDPGCGLIATVGHYWRQPIVADPLLHQCQCPQSGQVALIADCVQEPTRKCTFAFANGCL	127
DB	68	GLTDPGCGLIATVGHYWRQPIVADPLLHQCQCPQSGQVALIADCVQEPTRKCTFAFANGCL	128
QY	127	APQCALIVHTQADITSSDPDFNLKTCVCQGLINANVTADILEPAGGSIAFTGHYQAFR	186
DB	124	EPMA-----PGSPFAPWPIPVH-----TVPVAE-----	148
QY	187	PIFVNIFFQSNLC-----IKRQRQDSGSIQALISHLKVSADCVSLIAGVSL-----	245
DB	149	DKLRFPMRS-CPRYHELRE ACPAAHYQALQEWGTGTSKLENFTG-LSLVGPELR	203
QY	235	ASWLEIFLLQAGCMPPFCWGTETIDSHQWNTLLSLINACFELLQSTPEVARSATHLL	294
DB	204	PAWKVLDITMCQCAHGLPLPAAW-----SPWVPLTL--AQISALPILCAHVCPPEAKKA	255
QY	295	PI-----IKTALTPHPQKQAYGVPLTSVILFIAQHTVNIAMIGALEENHLLPGQHPINTP	350
DB	256	QITQCTGLNATLANFSPVQPIG--LPLKMYVSAHNTLLALACALGEL-----YACHTP	307



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen, Ltd.

OM protein - protein search, using sw model

Run on: December 13, 2002, 15:52:40 : Search time 21 seconds  
(without alignments)  
1995.934 Million cell updates/sec

Title: US-09-866-379A-10  
Perfect score: 2279  
Sequence: 1 MEALLFFKSLLELLRQSA..... : RPLQVMEALRFAELESSE 43C

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR\_23.\*

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution

# SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	2182	95.7	422	B36733	acid phosphatase (EC 3.1.3.2)
2	2144	94.1	434	B90776	phosphonhydrolase F
3	2144	94.1	444	B85633	hypothetical prote
4	959.5	42.1	441	A20201	acid phosphatase (EC 3.1.3.2)
5	560	24.6	413	B36773	periplasmic glucos
6	560	24.6	413	B85635	periplasmic glucos
7	554	24.3	413	JN0087	glucose-1-phosphat
8	547	24.0	413	A20534	glucose-1-phosphat
9	538	23.6	417	S25227	glucose-1-phosphat
10	536	23.5	414	B87316	periplasmic phosph
11	137	6.0	416	B16358	hypothetical prote
12	134.5	5.9	423	S86167	acid phosphatase (EC 3.1.3.2)
13	119	5.3	433	A23395	acid phosphatase (EC 3.1.3.2)
14	112.5	4.9	428	S64682	acid phosphatase (EC 3.1.3.2)
15	111.5	4.9	421	S14712	acid phosphatase (EC 3.1.3.2)
16	109.5	4.8	344	B87130	protein F52E1.8 (1)
17	109.5	4.8	457	A56925	paired box transcr
18	109.5	4.8	479	JN0890	acid phosphatase (EC 3.1.3.2)
19	108	4.7	396	JH0610	acid phosphatase (EC 3.1.3.2)
20	107.5	4.7	459	S52353	paired box transcr
21	107.5	4.7	479	JN0715	3-phytase (EC 3.1.3.1)
22	107	4.7	5126	S40450	tyrosine receptor
23	105	4.6	453	A54429	paired box transcr
24	104.5	4.6	397	C81716	hypothetical prote
25	104.5	4.5	537	S54770	secreted acid phos
26	104.5	4.5	888	S45726	secreted acid phos
27	104.5	4.5	1048	BV838C	exonuclease (EC 3.1.1.1)
28	103	4.5	769	B87681	tyrosine kinase Li
29	101.5	4.5	356	F15594	hypothetical prote

## ALIGNMENTS

### RESULT 1

B36733

acid phosphatase (EC 3.1.3.2) precursor - Escherichia coli (strain K-12)  
N.Altitude bases: phosphonhydrolase, periplasmic, pyruvate 6  
C.Species: Escherichia coli  
C.Date: 19-Apr-1991 #sequence\_revision 19-Apr-1991 #extl\_change 01-Mar-2002  
C.Accession: B36733; S18018; B64839; A26534; S17960; S33278  
R.Dassa, J.; Marck, C.; Boquet, P.L.  
J. Bacteriol. 172, 5497-5500, 1990  
A.Title: The complete nucleotide sequence of the Escherichia coli gene appA reveals  
A.Reference number: A36733; MUID:90368616; PMID:2168485

A.Accession: B36733

A.Status: preliminary

A.Molecule type: DNA

A.Residues: 1-432 <DAS>

A.Cross-references: ap M59708; NID:q145283; PIRN:AA72086.1; PIR:q145285

P.Greiner, P.; Jany, R.; D

Biol. Chem. Hoppe-Seyler 372, 664-665, 1991

A.Title: Characterization of a phytase from Escherichia coli.

A.Reference number: S18018

A.Accession: S18018

A.Molecule type: protein

A.Residues: 23-33 <DAS>

P.Plattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.

A. Rose, P.J.; Mau, R.; Shao, Y.

Science 277, 1453-1462, 1997

A.Title: The complete genome sequence of Escherichia coli K-12.

A.Reference number: A64726; MUID:97426617; PMID:9278503

A.Accession: B64839

A.Status: nucleic acid sequence not shown; translation not shown

A.Molecule type: DNA

A.Residues: 1-432 <BLAT>

A.Cross-references: GB:AE000200; GB:500296; NID:q2367111; PIRN:AA74065.1; PIR:q14717

A.Experimental Source: Strain K-12, substrain MG1655

P.Touati, F.; Danchin, A.

Biochimie 69, 215-221, 1997

A.Title: The structure of the promoter and amino terminal region of the prf 2.5 gene

A.Reference number: A26534; MUID:97271766; PMID:9032901

A.Accession: A26534

A.Molecule type: DNA

A.Residues: 1-50; NAGCHPRMANAG/65; T/67-74; DV/77-111; S/ <DAS>

A.Cross-references: GB:X05471; NID:q37925; PIRN:CAA24031.1; PIR:q404927

F.Dassa, J.; Fathi, H.; Marck, C.; Dion, M.; Kiefer-Bontemps, M.; Roquet, P.L.

Mol. Gen. Sect. 229, 341-352, 1991

A.Title: A new oxygen-regulated operon in Escherichia coli comprises the genes for

A.Reference number: S17958; MUID:92049231; PMID:1658595

A.Accession: S17958

A.Status: not compared with conceptual translation

A.Molecule type: DNA

A.Residues: 1-17 <DAS>

A.Cross-references: GB:S63811; NID:q28556; PIRN:AA620286.1; PIR:q238659

P.Greiner, F.; Kiefer-Bontemps, M.; Jany, R.D.

	Query Match	94.1%	Score: 2544	18	21	100000.4-42			
	Best local Similarity	96.5%	Prod. No. 1.10.163						
	Matches 419; Conservative	0;	Mismatches 13;	Indels	23	Gaps			
QY	1	MKALLPELSLLPLTPQSAPFAS	-	EPEKIKESVSVVVSQHVVEAT	IKAKLMLGVETPA	10			
IB	1	MKALLPELSLLPLTPQSAPFAS	EEPEKIKESVVEPEKIKESVSVVVSQHVVEAT	IKAKLMLGVETPA	10				
QY	59	WUTWVKRLDELTPGRIELIAY	ILYHWRQGVVALGELLK	YQ	YQSSAVAT	AMVETPK	128		
IB	61	WENIVSKWELTPGRIELIAY	ILYHWRQGVVALGELLK	YQ	YQSSAVAT	AMVETPK	12		
QY	119	GLALANLALGATLIVHQAQ	YQYQYQSSAVAT	YQ	YQSSAVAT	AMVETPK	128		
IB	121	GLALANLALGATLIVHQAQ	YQYQYQSSAVAT	YQ	YQSSAVAT	AMVETPK	128		
QY	179	QYVQIA	ELLLEEVNI	POSNNLEPEKIK	YQYQYQSSAVAT	YQ	YQSSAVAT	AMVETPK	208
IB	181	CHPQTATPELEEVNI	POSNNLEPEKIK	YQYQYQSSAVAT	YQ	YQSSAVAT	AMVETPK	240	

[illegible]

	Query Match	94.3%	Score 2144	Id: 27	Length 4443
	Host Local Similarity	96.5%	Prod. No. 1.1e+06		
	Matches 419	Conserved 396	0	Mismatches 18	Gaps 1
QY	1	MEATLPTSLTLPSPSAAGN	EPETKLESVWVSGHVAPEATKATLMGVVLA	98	
DB	1	TTTTTTTTTTTTTTTTTTTT	TTTTTTTTTTTTTTTTTTTT	100	
DB	11	MKAHTPLSLTLPQSAFAGSDI	EPETKLESVWVSGHVAPEATKATLMGVVLA	90	
QY	59	WPTMVALEGTTPGGEELAYLGYWPGPVALPFLKPTLSLVVAIAVLPFRK	118		
DB	71	WPNFVELSWLTPPGGEELAYLGYHYPGSPVAVGSLFRK	129	2	2
QY	119	GLAANLAFRTATVHTQATNSSTFETNPKTAWQENANVATLGGAGASIAET	178		
DB	181	GEATFAAGLAPCATVHTQATSSSDPFLNPKTAWQENANVATLSPKASIAET	190		

Db 366 ELWNIHNRKYVAVRMFYQIMQJAKNAKELKERRAGLISVAVAGNENKERRICELD 425  
 QY 417 GFTQVNEARIPACSL 432  
 Db 426 TFQKKVAKVIPACHI 441  
 RESULT 5  
 F90773  
 periplasmic glucose-1-phosphatase [imported] - Escherichia coli (strain 0157:H7, sub  
 C:Species: Escherichia coli  
 C:Date: 18-Jul-2001 #sequence\_revision 18 Jul 2001 #text\_change 18-Jul-2001  
 C:Accession: F90773  
 K:Hayashi, T.; Makino, K.; Ohnishi, M.; Katsukawa, K.; Ishii, K.; Yokoyama, K.; Han  
 gasawara, N.; Yasunaga, T.; Kikuta, S.; Shiba, T.; Hattori, M.; Shimadzu, H.  
 DNA Res. 8, 11-22, 2001  
 A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli 0157:H7 and  
 A:Reference number: A99629; M01021156231; PMID:11258796  
 A:Accession: F90773  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-413 <HAY>  
 A:Cross-references: GR-BA000007; PIDN:BA034581.1; PID:q13360618; GSPDB:GN00154  
 A:Experimental source: strain 0157:H7, substrain RIMP 0509952  
 C:Genetics:  
 A:Gene: ECs1158

Query Match 24.0%, Score 560, E5.2, Length 413;  
 Best Local Similarity 33.3%; Pred. No. 5.5e-37;  
 Matches 141; Conservative 73; Mismatches 174; Indels 36; Gaps 12;  
 QY 13 IFLETPQAFQSEPE-LKLESVIVSPQVAP-IFKATQIMQVTPVAMPVWKLGLF 70  
 Db 13 VYLLASNAQIVFEYGIQVLMSSRLNAPLANRHSVLEQSTPNKPEWPEVGRQIT 72  
 QY 71 PEGGELIAYCHYWPQPIVAPGHPKQCGQGVALLADVDERTKTCFAFAAGIAPDC 130  
 Db 73 IKGGLVEMVGMHREMLAFQGMVKSQGGEPDPIVAVANSQRTVAIAQFFIIGAFNP 132  
 QY 131 AITVITQADTSSPDPLENPLKIGVQLINANNVID--ALIEKAGKSIALFTGHYATAFE 187  
 Db 133 DIPVHGHQKMGTMDFNPIVD--DCAHSPQVAAVAKKELSKLGLHISYQI---- 184  
 QY 188 LERVINFPQSNLCLEKQDESCSTQALPSKLSYADCVSLIGAVSLASMLEIFLQQ 247  
 Db 184 LEKIVNYKDSFAC---REKQCSLVKGNFSAKVQGVKSVSGHLKVNLSLVDAFLQY 249  
 QY 248 AQGMP--PQWGRITTSQHWNTLLSLHNAQFDLQRTPEVARSRAFLIDLEKIALIDHP 305  
 Db 240 YEGFPMDQVAVGPIKSDQOMKVLKNGYQUSLFTSPVARNVAKPLVSYIDKAL---- 295  
 QY 306 PKQAYGVTLPTS---VLEFIAGHDTNLANGLALELN-WTIDQDQNTIPNGGELVPEWR 361  
 Db 296 -----VIDRISAPKILVILGHUSNLSLLTALGPKYQLHQHCKRTICKRTVPEWR 348  
 QY 392 KSLNEMZVSVFQTLQGMCKTLPESLNTFPCVKTLAGCFEENAVMSLAQFTQI 421  
 Db 349 DSRANDELKIEVYVSAGELKRNADALTQAFQGRVTFLESKP-FIDANSPFDMKPEUSV 407  
 QY 422 VNEA 425  
 Db 408 LNEA 411

RESULT 6  
 B85636  
 periplasmic glucose-1-phosphatase [imported] - Escherichia coli (strain 0157:H7, sub  
 C:Species: Escherichia coli  
 C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
 C:Accession: B85636  
 R:Perera, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mo  
 iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Ap  
 Nature 409, 529-533, 2001

QY 179 GHVQAFAPLEVLNFPQSNLCLEKQDESCSTQALPSKLSYADCVSLITGAVSLAS 238  
 Db 191 GHQVAFRELEVLNFPQSNLCLEKQDESCSTQALPSKLSYADCVSLITGAVSLAS 250  
 QY 239 LFEIFLQACAGMPPEWGRITDSQWNTLLSLHNAQFDLQRTPEVARSRAFLIDLIK 298  
 Db 251 LFEIFLQACAGMPPEWGRITDSQWNTLLSLHNAQFYIQTPEVAPSRATPLIDIM 310  
 QY 299 TALTHTPTQKQAYGVTLPSVLEFIAGHDTNLANGLALELNWTIDQDQNTIPPGGELVFE 358  
 Db 311 TALTHTPTQKQAYGVTLPSVLEFIAGHDTNLANGLALELNWTIDQDQNTIPPGGELVFE 370  
 QY 359 PWRPISDMSQVSVFQTLQGMCKTLPESLNTFPCVKTLAGCFEENAVMSLAQFTQI 421  
 Db 371 RWRPISDMSQVSVFQTLQGMCKTLPESLNTFPCVKTLAGCFEENAVMSLAQFTQI 430  
 QY 419 TQIVNEARIPACSL 432  
 Db 431 TQIVNEARIPACSL 444

RESULT 4  
 AC0201  
 acid phosphatase (EC 3.1.3.2) [imported] - Yersinia pestis (strain C092)  
 C:Species: Yersinia pestis  
 C:Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 27-Nov-2001  
 C:Accession: AC0201  
 K:Parkhill, J.; Wren, B.W.; Thomson, N.P.; Tittball, P.W.; Holden, M.T.G.; Prentice, M.B.;  
 deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;  
 H., M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,  
 Nature 413, 523-527, 2001  
 A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.  
 A:Reference number: AB0001; M01021470413; PMID:11586360  
 A:Accession: AC0201  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-441 <KOR>  
 A:Cross-references: GR-AL590842; PIDN:CA0290470.1; PID:q15976695; GSPDB:GN00175  
 C:Genetics:  
 A:Gene: YPO1648  
 C:Keywords: phosphoric monoester hydrolase

Query Match 42.1%; Score 959.5; DB 2; Length 441;  
 Best Local Similarity 46.1%; Pred. No. 6.5e-69;  
 Matches 201; Conservative 67; Mismatches 153; Indels 15; Gaps 7;  
 QY 5 LIPFLSLIDPLTPQSAFAQSEPE-LKLESVIVSPQVAP-IFKATQIMQVTPVAMPVW 63  
 Db 13 LVLMISGLAATAVA---AFPSGYTLERVVILSRHGVSRSPTKOTLMNDVTPDKMPQW 69  
 QY 64 VKGSELTPRGELIAYCHYWPQPIVAPGHPKQCGQGVALLADVDERTKTCFAFA 123  
 Db 70 VKAGVLTPRGAEVTLMGGFYGYDFRSIGLL-AAGCPARGVVAQADIDQFTLTCGAFI 128  
 QY 124 ACLADKATVHTQADTSSPDPLENPLKIGVQLINANNVIDALIERAGGSIAUFIQHYOT 183  
 Db 129 DGVAPGGLTVINQADLIKTDPLPHVYAGVCKLDAQTDKAIEQLGGPLDVTVSQYAK 188  
 QY 184 AFRELEVLNFPQSNLCLEKQDESCSTQALPSKLSYADCVSLITGAVSLASMLEITE 241  
 Db 189 DPAQMGVNLNFAASPYCKLQOQKTCDFAHFAANEVNNKKGKVTLSGPIALSSTIGE 248  
 QY 242 LFLQACAGMPPEWGRITDSQWNTLLSLHNAQFDLQRTPEVARSRAFLIDLIK 301  
 Db 249 LFLQACAGMPPEWGRITDSQWNTLLSLHNAQFYIQTPEVAPSRATPLIDIM 308  
 QY 302 THTHTPTQKQAYGVTLPTS---VLEFIAGHDTNLANGLALELNWTIDQDQNTIPPGGELV 357  
 Db 309 T---LQIDAQAGKLPISAQNKVILFIAGHDTNLANGLALELNWTIDQDQNTIPPGGELV 365  
 QY 458 EWRPPLSDMSQVSVFQTLQGMCKTLPESLNTFPCVKTLAGCFEENAVMSLA 416

C:Genetics:  
A:Gene: *acp*  
A:Map position: 23 min  
C:Function:  
A:Description: essential for growth in a high phosphate medium containing glucose-1-phosphate  
A:Note: optimal at low pH  
A:Key: 1, immediate; 2, periplasmic space; phosphatidyl; phosphatidyl; phosphatidyl; phosphatidyl  
F:1-22/Domain: signal sequence #status predicted - NTC  
F:23-413/Product: glucose-1-phosphate #status predicted - MAL  
F:40/Active site: His (phosphatidyl; intermediate) #status predicted  
F:311/Active site: His #status predicted

[illegible]

RESULT 8  
 AG0642  
 glucose-1-phosphatase precursor (clpase), secreted (imported) Salmonella enterica  
 A:Species: Salmonella enterica subsp. enterica serovar typhi  
 A:Name: AG0642  
 A:Date: 04-Nov-2001  
 A:Accession: AG0642  
 A:Accession: AG0642  
 A:Author: J. J. Bourdau, G. J. James, K. D. J. Thomson, N. K. J. Pirkard, J. J. White, J. J. Chou  
 A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica  
 A:Reference number: AB0502; PMID:11677608  
 A:Accession: AG0642  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1413 rPAR  
 A:Cross-references: GB:AL514462; PIDN:AP004241; PIR:J052004; Q84456N  
 C:Genetics:  
 A:Gene: STY1163  
 Query Match 24.0% Score 547 Length 414  
 Query Local Similarity 34.7% Pred. No. 6046





## RESULT 11

T16058

Hypothetical protein F13D11.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 23-Mar-2001

C:Accession: T16058

R:Fullon, L.

submitted to the EMBL Data Library, November 1995

A:Description: The sequence of C. elegans cosmid F13D11.

A:Reference number: S69020

A:Accession: T16058

A&gt;Status: preliminary, translated from cDNA exp. gene

A:Molecule type: DNA

A:Residues: 1-416 sFHLs

A:Cross-references: EMBL:U49049; NID:q1072175; PID:q1072175; FID:AAA81702.1; CESP:F13D11

C:Genetics:

A:Gene: CESP:F13D11.1

A:Introns: 2; exons 9; 963, 1243, 1277, 1443, 2763, 4127, 4473, 4773

C:Superfamily: mammalian acid phosphatase

Query Match 6.0%; Score 137; DB 2; Length 416;

Best Local Similarity 21.7%; Pred. No. 0.0049;

Matches 99; Conservation 61; Mismatch 166; Indels 130; Gaps 23;

QY 1 MKAILLPLSLPLPTPSAQAQSPDEKISVIVSVSHVAPTKATQLMGVTPD--- 57

DB 1 MKLVILLE-FLFW---AEV---KLKPVQIWERGERTPCHYLYFPCDGLNNVY 48

QY 58 AWITWIKVIGLPTPEPFTLFTSHYWEKLI-----VAIHLKRCSTYSSQVAI 107

DB 49 QQLAMP---GELTKRGLIFPPQR---QPIPTVPHPCVIVYQPP-----DEIV 91

QY 108 LAUWIERTKTEAFAGLAIFCALITVHIDALTNSSPDLFPLKRTGVQLDMANVTDAII 167

DB 92 YTKGNRTSASAGAMPAGFLPP---NEQJNNYELKQIV-----ACITDISI 146

QY 168 FRAMASIAIDFTCHYQTAFRE-----LQVINTFQSNILKRRKODSSCSUTQALPSIL 220

DB 147 DWVSIGALDNCVYGEAQKRSXAEVMDQEKVIAELQLQVRNIADP----- 184

QY 221 KVAIAVSIIGAVSIASMTETPTTGGAGCMPEPQWGLTTSHEWNILLSHRAQV 277

DB 185 -----PIVAVKYNHVIDSKVRYILQDQR-LIYPPKACYENRIINMSFLHDA--- 233

QY 278 ILQTPPEVAPATPIIDN-----ITATLTPHPAFQAYGVLTFTSVLFIAGHCHLAHLG 333

DB 234 -----VKKVGNISVGVNHFVMSYFTHLQKNSIKGV-----FISCHDNLVITW 279

QY 334 GALEFNWTLPGQIPNTPP-GGELVFERWRKRLSDNSQWLVQSVLVPGTLQMPKRTPLSLNT 392

DB 280 ESLRD-----GHPDIPNYGAHIAE-----MHFVGLSLTKFPLSMGP 319

QY 393 PIGAVKELTACGERNAQCM-SLAETFTLVNIAETP 428

DB 420 NCTRVELHPHC-SKSNNDCTWDEPQRVKKSKPK 454

## RESULT 12

S06167

acid phosphatase (E0 3.1.3.2) ACT2 precursor [validated] - human

N:Alternate names: acid phosphatase, lysosomal

C:Species: Homo sapiens (man)

C:Date: 04-Dec-1992 #sequence\_revision 01-Dec-1995 #text\_change 08-Dec-2000

C:Accession: S06167; S05525; S01155

R:von Figura, R.

submitted to the EMBL Data Library, June 1989

A:Reference number: S06167

A:Accession: S06167

A:Molecule type: DNA

A:Residues: 1-423 vONS

A:Cross-references: EMBL:X15525; NID:q44249; FID:CAA3542.1; PID:q1194524

R:Geiger, C.; von Figura, R.

Eur. J. Biochem. 183, 611-616, 1989

A:Title: Structure of the human lysosomal acid phosphatase gene

A:Reference number: S05525; MIM:894792.8; PMID:276354

A:Accession: S05525

A:Molecule type: DNA

A:Residues: 1-29 vE15

R:Pohlmann, R.; Krentler, C.; Schmidt, R.; Schroeder, W.; Lorkowski, G.; Calley, I.

EMBO J. 7, 2343-2350, 1988

A:Title: Human lysosomal acid phosphatase: cloning, expression and chromosomal assignment

A:Reference number: S01155; MIM:890524.5; PMID:319103

A:Accession: S01155

A:Molecule type: mRNA

A:Residues: 1-423 vO8

A:Cross-references: EMBL:X12548; NID:q44262; FID:CAA31064.1; PMID:61204

A&gt;Note: part of this sequence, including the amino end of the mature protein, was assigned to the mature protein

C:Genetics:

A:Gene: GDB:ACT2

A:Cross-references: GDB:118663; OMIM:171650

A:M4 position: 17111-21111

A:Introns: 6; exons 9; 603, 923, 1703, 1863, 2103, 2203, 2303, 2403, 2503, 2603, 2703, 2803, 2903, 3003, 3103, 3203, 3303, 3403, 3503, 3603, 3703, 3803, 3903, 4003, 4103, 4203, 4303, 4403, 4503, 4603, 4703, 4803, 4903, 5003, 5103, 5203, 5303, 5403, 5503, 5603, 5703, 5803, 5903, 6003, 6103, 6203, 6303, 6403, 6503, 6603, 6703, 6803, 6903, 7003, 7103, 7203, 7303, 7403, 7503, 7603, 7703, 7803, 7903, 8003, 8103, 8203, 8303, 8403, 8503, 8603, 8703, 8803, 8903, 9003, 9103, 9203, 9303, 9403, 9503, 9603, 9703, 9803, 9903, 10003, 10103, 10203, 10303, 10403, 10503, 10603, 10703, 10803, 10903, 11003, 11103, 11203, 11303, 11403, 11503, 11603, 11703, 11803, 11903, 12003, 12103, 12203, 12303, 12403, 12503, 12603, 12703, 12803, 12903, 13003, 13103, 13203, 13303, 13403, 13503, 13603, 13703, 13803, 13903, 14003, 14103, 14203, 14303, 14403, 14503, 14603, 14703, 14803, 14903, 15003, 15103, 15203, 15303, 15403, 15503, 15603, 15703, 15803, 15903, 16003, 16103, 16203, 16303, 16403, 16503, 16603, 16703, 16803, 16903, 17003, 17103, 17203, 17303, 17403, 17503, 17603, 17703, 17803, 17903, 18003, 18103, 18203, 18303, 18403, 18503, 18603, 18703, 18803, 18903, 19003, 19103, 19203, 19303, 19403, 19503, 19603, 19703, 19803, 19903, 20003, 20103, 20203, 20303, 20403, 20503, 20603, 20703, 20803, 20903, 21003, 21103, 21203, 21303, 21403, 21503, 21603, 21703, 21803, 21903, 22003, 22103, 22203, 22303, 22403, 22503, 22603, 22703, 22803, 22903, 23003, 23103, 23203, 23303, 23403, 23503, 23603, 23703, 23803, 23903, 24003, 24103, 24203, 24303, 24403, 24503, 24603, 24703, 24803, 24903, 25003, 25103, 25203, 25303, 25403, 25503, 25603, 25703, 25803, 25903, 26003, 26103, 26203, 26303, 26403, 26503, 26603, 26703, 26803, 26903, 27003, 27103, 27203, 27303, 27403, 27503, 27603, 27703, 27803, 27903, 28003, 28103, 28203, 28303, 28403, 28503, 28603, 28703, 28803, 28903, 29003, 29103, 29203, 29303, 29403, 29503, 29603, 29703, 29803, 29903, 30003, 30103, 30203, 30303, 30403, 30503, 30603, 30703, 30803, 30903, 31003, 31103, 31203, 31303, 31403, 31503, 31603, 31703, 31803, 31903, 32003, 32103, 32203, 32303, 32403, 32503, 32603, 32703, 32803, 32903, 33003, 33103, 33203, 33303, 33403, 33503, 33603, 33703, 33803, 33903, 34003, 34103, 34203, 34303, 34403, 34503, 34603, 34703, 34803, 34903, 35003, 35103, 35203, 35303, 35403, 35503, 35603, 35703, 35803, 35903, 36003, 36103, 36203, 36303, 36403, 36503, 36603, 36703, 36803, 36903, 37003, 37103, 37203, 37303, 37403, 37503, 37603, 37703, 37803, 37903, 38003, 38103, 38203, 38303, 38403, 38503, 38603, 38703, 38803, 38903, 39003, 39103, 39203, 39303, 39403, 39503, 39603, 39703, 39803, 39903, 40003, 40103, 40203, 40303, 40403, 40503, 40603, 40703, 40803, 40903, 41003, 41103, 41203, 41303, 41403, 41503, 41603, 41703, 41803, 41903, 42003, 42103, 42203, 42303, 42403, 42503, 42603, 42703, 42803, 42903, 43003, 43103, 43203, 43303, 43403, 43503, 43603, 43703, 43803, 43903, 44003, 44103, 44203, 44303, 44403, 44503, 44603, 44703, 44803, 44903, 45003, 45103, 45203, 45303, 45403, 45503, 45603, 45703, 45803, 45903, 46003, 46103, 46203, 46303, 46403, 46503, 46603, 46703, 46803, 46903, 47003, 47103, 47203, 47303, 47403, 47503, 47603, 47703, 47803, 47903, 48003, 48103, 48203, 48303, 48403, 48503, 48603, 48703, 48803, 48903, 49003, 49103, 49203, 49303, 49403, 49503, 49603, 49703, 49803, 49903, 50003, 50103, 50203, 50303, 50403, 50503, 50603, 50703, 50803, 50903, 51003, 51103, 51203, 51303, 51403, 51503, 51603, 51703, 51803, 51903, 52003, 52103, 52203, 52303, 52403, 52503, 52603, 52703, 52803, 52903, 53003, 53103, 53203, 53303, 53403, 53503, 53603, 53703, 53803, 53903, 54003, 54103, 54203, 54303, 54403, 54503, 54603, 54703, 54803, 54903, 55003, 55103, 55203, 55303, 55403, 55503, 55603, 55703, 55803, 55903, 56003, 56103, 56203, 56303, 56403, 56503, 56603, 56703, 56803, 56903, 57003, 57103, 57203, 57303, 57403, 57503, 57603, 57703, 57803, 57903, 58003, 58103, 58203, 58303, 58403, 58503, 58603, 58703, 58803, 58903, 59003, 59103, 59203, 59303, 59403, 59503, 59603, 59703, 59803, 59903, 60003, 60103, 60203, 60303, 60403, 60503, 60603, 60703, 60803, 60903, 61003, 61103, 61203, 61303, 61403, 61503, 61603, 61703, 61803, 61903, 62003, 62103, 62203, 62303, 62403, 62503, 62603, 62703, 62803, 62903, 63003, 63103, 63203, 63303, 63403, 63503, 63603, 63703, 63803, 63903, 64003, 64103, 64203, 64303, 64403, 64503, 64603, 64703, 64803, 64903, 65003, 65103, 65203, 65303, 65403, 65503, 65603, 65703, 65803, 65903, 66003, 66103, 66203, 66303, 66403, 66503, 66603, 66703, 66803, 66903, 67003, 67103, 67203, 67303, 67403, 67503, 67603, 67703, 67803, 67903, 68003, 68103, 68203, 68303, 68403, 68503, 68603, 68703, 68803, 68903, 69003, 69103, 69203, 69303, 69403, 69503, 69603, 69703, 69803, 69903, 70003, 70103, 70203, 70303, 70403, 70503, 70603, 70703, 70803, 70903, 71003, 71103, 71203, 71303, 71403, 71503, 71603, 71703, 71803, 71903, 72003, 72103, 72203, 72303, 72403, 72503, 72603, 72703, 72803, 72903, 73003, 73103, 73203, 73303, 73403, 73503, 73603, 73703, 73803, 73903, 74003, 74103, 74203, 74303, 74403, 74503, 74603, 74703, 74803, 74903, 75003, 75103, 75203, 75303, 75403, 75503, 75603, 75703, 75803, 75903, 76003, 76103, 76203, 76303, 76403, 76503, 76603, 76703, 76803, 76903, 77003, 77103, 77203, 77303, 77403, 77503, 77603, 77703, 77803, 77903, 78003, 78103, 78203, 78303, 78403, 78503, 78603, 78703, 78803, 78903, 79003, 79103, 79203, 79303, 79403, 79503, 79603, 79703, 79803, 79903, 80003, 80103, 80203, 80303, 80403, 80503, 80603, 80703, 80803, 80903, 81003, 81103, 81203, 81303, 81403, 81503, 81603, 81703, 81803, 81903, 82003, 82103, 82203, 82303, 82403, 82503, 82603, 82703, 82803, 82903, 83003, 83103, 83203, 83303, 83403, 83503, 83603, 83703, 83803, 83903, 84003, 84103, 84203, 84303, 84403, 84503, 84603, 84703, 84803, 84903, 85003, 85103, 85203, 85303, 85403, 85503, 85603, 85703, 85803, 85903, 86003, 86103, 86203, 86303, 86403, 86503, 86603, 86703, 86803, 86903, 87003, 87103, 87203, 87303, 87403, 87503, 87603, 87703, 87803, 87903, 88003, 88103, 88203, 88303, 88403, 88503, 88603, 88703, 88803, 88903, 89003, 89103, 89203, 89303, 89403, 89503, 89603, 89703, 89803, 89903, 90003, 90103, 90203, 90303, 90403, 90503, 90603, 90703, 90803, 90903, 91003, 91103, 91203, 91303, 91403, 91503, 91603, 91703, 91803, 91903, 92003, 92103, 92203, 92303, 92403, 92503, 92603, 92703, 92803, 92903, 93003, 93103, 93203, 93303, 93403, 93503, 93603, 93703, 93803, 93903, 94003, 94103, 94203, 94303, 94403, 94503, 94603, 94703, 94803, 94903, 95003, 95103, 95203, 95303, 95403, 95503, 95603, 95703, 95803, 95903, 96003, 96103, 96203, 96303, 96403, 96503, 96603, 96703, 96803, 96903, 97003, 97103, 97203, 97303, 97403, 97503, 97603, 97703, 97803, 97903, 98003, 98103, 98203, 98303, 98403, 98503, 98603, 98703, 98803, 98903, 99003, 99103, 99203, 99303, 99403, 99503, 99603, 99703, 99803, 99903, 100003, 100013, 100023, 100033, 100043, 100053, 100063, 100073, 100083, 100093, 100103, 100113, 100123, 100133, 100143, 100153, 100163, 100173, 100183, 100193, 100203, 100213, 100223, 100233, 100243, 100253, 100263, 100273, 100283, 100293, 100303, 100313, 100323, 100333, 100343, 100353, 100363, 100373, 100383, 100393, 100403, 100413, 100423, 100433, 100443, 100453, 100463, 100473, 100483, 100493, 100503, 100513, 100523, 100533, 100543, 100553, 100563, 100573, 100583, 100593, 100603, 100613, 100623, 100633, 100643, 100653, 100663, 100673, 100683, 100693, 100703, 100713, 100723, 100733, 100743, 100753, 100763, 100773, 100783, 100793, 100803, 100813, 100823, 100833, 100843, 100853, 100863, 100873, 100883, 100893, 100903, 100913, 100923, 100933, 100943, 100953, 100963, 100973, 100983, 100993, 101003, 101013, 101023, 101033, 101043, 101053, 101063, 101073, 101083, 101093, 101103, 101113, 101123, 101133, 101143, 101153, 101163, 101173, 101183, 101193, 101203, 101213, 101223, 101233, 101243, 101253, 101263, 101273, 101283, 101293, 101303, 101313, 101323, 101333, 101343, 101353, 101363, 101373, 101383, 101393, 101403, 101413, 101423, 101433, 101443, 101453, 101463, 101473, 101483, 101493, 101503, 101513, 101523, 101533, 101543, 101553, 101563, 101573, 101583, 101593, 101603, 101613, 101623, 101633, 101643, 101653, 101663, 101673, 101683, 101693, 101703, 101713, 101723, 101733, 101743, 101753, 101763, 101773, 101783, 101793, 101803, 101813, 101823, 101833, 101843, 101853, 101863, 101873, 101883, 101893, 101903, 101913, 101923, 101933, 101943, 101953, 101963, 101973, 101983, 101993, 102003, 102013, 102023, 102033, 102043, 102053, 102063, 102

C:Accession: A33395  
R:Himeno, M.; Fujita, H.; Nomuchi, Y.; Kono, A.; Kato, K.  
Biochem. Biophys. Res. Commun. 162, 1044-1053, 1989  
A:Title: Isolation and sequencing of a cDNA clone encoding acid phosphatase in rat liver  
A:Reference number: A33395; MUID:89250910; PMID:2764916  
A:Accession: A33395  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-423 <CHIM>  
A:Cross-references: GB:827693, EMBL:944433, F001 AAA12714 1; F001 9302934  
C:Superfamily: mammalian acid phosphatase  
C:Keywords: phosphatidylidene phosphatase  
F:41/Active site: Arg #status predicted  
F:42/Active site: His (phosphohistidine intermediate) #status predicted

Query Match 5.2%; Score 119.0; DB 2; Length 423;  
Best Local Similarity 22.9%; Pred. No. 11;  
Matches 105; Conservative 46; Mismatches 188; Indels 120; Gaps 23;

QY 2 KAILIPPL--SLIPLTPSAFAQSEPELKLESVIVSRHGVRAPIKATQMLMDVTPDAM 59  
DB 10 QAAALQFLQMLQVWPPIQA-----RSLRFVTLRYRGDRSPVKA-----YKDPYQE 58  
QY 60 PTWPVKLGELTPGGELIAYLGHYWRQRLVADGLLPKCGPQSGQVAIIADVD-ETRTKT 118  
DB 59 EKWPQGGQLTKKGMQLHQLHQLGALRQY--HGFL--NASHYRQEVYVFSTDFDILMS 113  
QY 119 GEAFAGLADPCATVHTQADTSSPDPLFNPLKTGVQCLDNANVTDAILERAGGSTADPT 178  
DB 114 AENLAGLIPP-----TEQHFNPNSNOPI-----PVHTVPIE----- 148  
QY 179 GHYQTAFLERLVNTPQSMICLREKQDESCSLTQALFSELKVSACVSLTGAVSLASM 238  
DB 149 -----DRLIKFPLGHCPRYEQAGNE-----TRQIPYQMSIQNAQFLDMVANEIC 194  
QY 239 LTELFL-----LQACGMPEPGWGRITDSHOWNTLLSLHNAQF-----DLLO 280  
DB 195 LNNLTLETWNVYDTLECRTHGLLLPPW---ASPTQVQALSQLKDFSLFLRGHDOVQ 251  
QY 281 PTFVAPSPAPVPIIDILKTATPHDPQKQAYGVTVIPSVIFACHDTNINIGGALINW 340  
DB 252 K-----APLQAVIQAQIKNIITIMATTSQF-----PKLIYSAHDTTLVALOMLNVY- 300  
QY 341 TLPCQDPTPGGELVFERRRSLDSQWLVQSLVFLQCMKDKTFLSLNTPGGEVKLT 400  
DB 301 -----NGKAPVASYHIFETVQ--KDNQNF--SVEMYPNDISK---KAPW-----PLT 341  
QY 401 LAGFEERNAQGMCSLAGTQIVNEARIP-----ACSLRS 434  
DB 342 LPGGPHR- ---CPLQDFLRL-TEPVTPKDWKRECLAS 374

RESULT 14  
S64682  
acid phosphatase (EC 3.1.3.2) Arph 1 precursor, lysosomal - fruit fly (Drosophila melanogaster)  
C:Species: Drosophila melanogaster  
C:Date: 06 Dec 1996 #sequence\_revision: 66-Dec-1996 #text\_change: 17 Mar-2006  
C:Accession: S64682; S64681  
R:Chung, H.J.; Shaffer, C.; MacIntyre, R.  
Mol. Gen. Genet. 250, 645-646, 1996  
A:Title: Molecular characterization of the lysosomal acid phosphatase from Drosophila melanogaster  
A:Reference number: S64681; MUID:96194627; PMID:8576866  
A:Accession: S64682  
A>Status: not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-438 <CHIM>  
A:Accession: S64681  
A>Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-31, 'Y', 33-438 <CHIM>  
C:Genot loc:  
A:Gene: Flybase:Arph-1  
A:Cross-references: Flybase:FBan0000042

C:Superfamily: mammalian acid phosphatase  
C:Keywords: phosphoric monoester hydrolase  
F:1-33/Domain: signal sequence #status predicted sSIG  
F:34-436/Product: acid phosphatase Arph-1 #status predicted -MAI-

Query Match 4.9%; Score 112.5; DB 2; Length 448;  
Best Local Similarity 20.6%; Pred. No. 0.38;  
Matches 82; Conservative 61; Mismatches 147; Indels 109; Gaps 19;

QY 29 FLESVIVSRHGVRAPIKATQIMQAVTTPDAMPT-----WPVKLGELTPGGELIAYL 80  
DB 47 GKRVHVIYFHSRTEFV --- --DAYTFQWGLSEHFWTQWENFNLPQEHYDL 95  
QY 81 GHYWRQRLVADGLLPKCGPQSGQVAIIAEVLEEFETSEAFASGL- AFPC----- 130  
DB 96 GKWLFPNY--SNLLPPI---YSNENIYVQSTDVDTLMSAQSNIAAGYFQCGEDWINDI 150  
QY 131 ---ATVITVQAUTSSPDPLFNPLKTGVQCLDNANVTDAILERAGGSTADPTSHYVTAFFE 187  
DB 151 NWQPIPIHTSPERDP-----ILAAKAPCPAYDY-----E 180  
QY 188 LERLVNFPQSMICLREKQDESCSLTQALFSELKVSACVSLTGAVSLASMLTFLFLLQ 247  
DB 181 LASLESPEFK-ALTEKHRNLFAYLSEKGGRPVKTFID-----AQVLNNTLFLEN 229  
QY 248 AGMPEPGWGRITDSHOWNTLLSLHNAQFDLLQRTPEVARSRATPLL-DLIKIALTPHP 306  
DB 230 LYNNMLLAKTKKVVGREELTVYS--NFAPAISSTYKLAHLKACPLFLVDFQRF----- 281  
QY 307 QKQAYGVTLPTSVLPI-AGHDNTNLANGLALNWLPAQDNTPEPGGELVFERRRSLD 365  
DB 282 KEKSSGSLKPDKRSMMVYSAHDTTVASVLNALKL-FEL-----HSPYFACIMMELR--VD 334  
QY 365 NSLWLVQSLVFLQCMKDKTFLSLNTPGGEVKLTLAG 404  
DB 334 ETNTELVSTFYK-----NTTAEPLPLDIPG 359

RESULT 15  
S14742  
acid phosphatase (EC 3.1.3.2) precursor - mouse (Mus musculus)  
C:Species: Mus musculus (house mouse)  
C:Date: 21 Nov 1993 #sequence\_revision: 10-Nov-1995 #text\_change: 31 Mar-2000  
C:Accession: S14742  
R:Geier, C.; von Figura, K.; Pohlmann, R.  
Biochem. Biophys. Res. Commun. 162, 301-304, 1989  
A:Title: Molecular cloning of the mouse lysosomal acid phosphatase  
A:Reference number: S14742; MUID:91282986; PMID:2059337  
A:Accession: S14742  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-421 <CEIF>  
A:Cross-references: DMBL:X57199; NID:952979; PIDN:CAA040485.1; PID:952871  
C:Superfamily: mammalian acid phosphatase  
C:Keywords: lysosoma; phosphoric monoester hydrolase

Query Match 4.9%; Score 111.5; DB 2; Length 421;  
Best Local Similarity 21.4%; Pred. No. 0.43;  
Matches 98; Conservative 53; Mismatches 171; Indels 147; Gaps 23;

QY 2 KAILIPPL--SLIPLTPSAFAQSEPELKLESVIVSRHGVRAPIKATQMLMDVTPDAM 56  
DB 8 QAAALQFLQMLQVWPPIQA-----RSLRFVTLRYRGDRSPVKA-----YKDPYQE 54  
QY 57 DAMPTWPVKLGELTPGGELIAYLGHYWRQRLVADGLLPKCGPQSGQVAIIADVD-ETRTKT 115  
DB 54 YQEEKWPQGGQLTKKGMQLHQLHQLGALRQY--HGFL--NASHYRQEVYVFSTDFDILMS 108  
QY 116 RKTGEAFAGLADPCATVHTQADTSSPDPLFNPLKTGVQCLDNANVTDAILERAGGSTA 175  
DB 109 LMSAEANLAGLIPP-----NEVOHSPNLSWQPI-----PVHTVPIE----- 146  
QY 176 DFTGHVQTAFFELFVNTFQSNICLPFFQDPSQSTTQALDPLKVSACVSLTGAVSL 235



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 C. Murphy, et al

OM protein - protein search, using sw model

Run on: December 13, 2002, 15:49:35, Search time 40 Seconds  
(without alignments)  
1452.132 Million cell updates/sec

Title: US-09-866-379A-10

Perfect score: 2279

Sequence: 1 MKAILFFLLPLPQEA.....GFTGVNEAFIPNCSLESHL 436

Scoring table: RIGSUM62

Gapop 10.0, Gapext 0.5

Searched: 908470 seqs, 13320620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2309030306

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :	A_Geneseq_101002.*
1:	/SILS2/seqdata/geneseq/geneseq-emb1/AA1940.DAT*
2:	/SILS2/seqdata/geneseq/geneseq-emb1/AA1981.DAT*
3:	/SILS2/seqdata/geneseq/geneseq-emb1/AA1982.DAT*
4:	/SILS2/seqdata/geneseq/geneseq-emb1/AA1983.DAT*
5:	/SILS2/seqdata/geneseq/geneseq-emb1/AA1984.DAT*
6:	/SILS2/seqdata/geneseq/geneseq-emb1/AA1985.DAT*
7:	/SILS2/seqdata/geneseq/geneseq-emb1/AA1986.DAT*
8:	/SILS2/seqdata/geneseq/geneseq-emb1/AA1987.DAT*
9:	/SILS2/seqdata/geneseq/geneseq-emb1/AA1988.DAT*
10:	/SILS2/seqdata/geneseq/geneseq-emb1/AA1989.DAT*
11:	/SILS2/seqdata/geneseq/geneseq-emb1/AA1990.DAT*
12:	/SILS2/seqdata/geneseq/geneseq-emb1/AA1991.DAT*
13:	/SILS2/seqdata/geneseq/geneseq-emb1/AA1992.DAT*
14:	/SILS2/seqdata/geneseq/geneseq-emb1/AA1993.DAT*
15:	/SILS2/seqdata/geneseq/geneseq-emb1/AA1994.DAT*
16:	/SILS2/seqdata/geneseq/geneseq-emb1/AA1995.DAT*
17:	/SILS2/seqdata/geneseq/geneseq-emb1/AA1996.DAT*
18:	/SILS2/seqdata/geneseq/geneseq-emb1/AA1997.DAT*
19:	/SILS2/seqdata/geneseq/geneseq-emb1/AA1998.DAT*
20:	/SILS2/seqdata/geneseq/geneseq-emb1/AA1999.DAT*
21:	/SILS2/seqdata/geneseq/geneseq-emb1/AA2000.DAT*
22:	/SILS2/seqdata/geneseq/geneseq-emb1/AA2001.DAT*
23:	/SILS2/seqdata/geneseq/geneseq-emb1/AA2002.DAT*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2282	97.8	430	AAE15808	Escherichia coli a
2	2188	96.0	440	AAE15809	Escherichia coli a
3	2188	96.0	440	AAE22635	Escherichia coli p
4	2188	96.0	440	AAE15806	Escherichia coli B
5	2182	95.7	432	AAU77775	Phytase associated
6	2182	95.7	432	AAE36257	Lama2/APPA plasmid
7	2182	95.7	432	AAE36258	R15/APPA plasmid t
8	2182	95.7	432	AAE36259	P15/APPA plasmid t
9	2182	95.7	432	AAE36260	P15/APPA plasmid t
10	2182	95.7	432	AAE36261	P15/APPA plasmid t

11	2182	95.7	442	AAE36262	SV40/APPA plasmid
12	2182	95.7	442	AAE36263	Escherichia coli a
13	2182	95.7	442	AAE15807	Escherichia coli a
14	2177	95.5	432	AAE02631	E. coli acid phosph
15	2171	95.3	432	AAE02634	E. coli acid phosph
16	2170	95.2	432	AAE02635	E. coli acid phosph
17	2157	94.6	432	AAE02632	E. coli acid phosph
18	2147	94.2	432	AAE34753	E. coli acid phosph
19	2080.5	91.3	423	AAE01513	Am Escherichia coli
20	1192	52.3	261	22 APG35698	Novel human diadmo
21	637	28.0	144	22 APG34256	Novel human diadmo
22	544	23.9	123	22 APG3269	Novel human enzyme
23	318	13.6	118	22 APG35694	Novel human diadmo
24	291	8.8	426	22 APG33348	Novel human diadmo
25	291	8.8	426	22 APG35788	Human protein phos
26	231	8.8	426	22 APG39880	Novel human secret
27	184.5	8.1	1062	22 APG19301	Novel human diadmo
28	131.5	5.8	423	22 APG14067	Human lysosomal ac
29	127	5.6	395	22 APG22558	Drosophila melanog
30	126	5.5	440	21 APG69549	Aspergillus fumiga
31	126	5.5	465	19 APG84356	Aspergillus fumiga
32	126	5.5	465	20 APG99005	A. fumigatus phyta
33	125	5.5	440	21 APG20507	Aspergillus fumiga
34	125	5.5	440	21 APG69550	Aspergillus fumiga
35	125	5.5	467	22 APG02117	A. fumigatus 14074
36	124	5.4	467	20 APG43171	A. fumigatus phyta
37	124	5.4	467	21 APG70528	Aspergillus fumiga
38	124	5.4	467	21 APG69570	Mutant Aspergillus
39	124	5.4	467	22 APG02112	A. fumigatus 14074
40	123.5	5.4	392	22 APG1760	Drosophila melanog
41	122	5.4	467	21 APG69574	Mutant Aspergillus
42	118	5.2	440	21 APG20508	Aspergillus fumiga
43	118	5.2	440	21 APG69551	Aspergillus fumiga
44	117	5.1	582	22 APG21719	Novel human diadmo
45	116.5	5.1	438	22 APG58803	Drosophila melanog

# ALIGNMENTS

RESULT 1	AAE15808
IL	AAE15806 Standard; Protein, 430 AA.
XX	AAE15808;
XX	26 MAR 2002 (first entry)
DT	Escherichia coli appA phytase mutant protein.
XX	
XX	Bacterial phytase; K12 appA phytase, phytase stability, analcolic
XX	gastrointestinal, nutritional value, feed treatment, phytase therapy
KW	thermal tolerance, growth performance, alcohol in drink, biopolymers
KW	non-alcoholic drink; bioleaching; mutant; mutin.
XX	Escherichia coli.
OS	Synthetic
XX	
PH	Key Location/Qualifiers
FT	Misc-difference 68 /note= "Wild type Trp substituted with Glu"
FT	Misc-difference 84 /note= "Wild type Glu substituted with Trp"
FT	Misc-difference 95 /note= "Wild type Ala substituted with Pro"
FT	Misc-difference 97 /note= "Wild type Lys substituted with Cys"
FT	Misc-difference 168 /note= "Wild type Ser substituted with Glu"
FT	Misc-difference 180 /note= "Wild type Arg substituted with Tyr"
FT	Misc-difference 225 /note= "Wild type Asn substituted with Cys"



DB 21: ETEFLQQAQGMPEGVVPTSVIETGSHQWSTETETIRACFYLCTETPEVAFSEATETLDTMAA 360  
 QY 301 LTPHPPOKQAYGVTLPTSVIETAGHDTNLANIGGALFNWTLPGGPTNTPGGSEIVFERW 360  
 DB 301 LTPHPPOKQAYGVTLPTSVIETAGHDTNLANIGGALFNWTLPGGPTNTPGGSEIVFERW 360  
 QY 361 PPSNSQWIAVSVIETGSHQWSTETETIRACFYLCTETPEVAFSEATETLDTMAA 420  
 DB 361 PPSNSQWIAVSVIETGSHQWSTETETIRACFYLCTETPEVAFSEATETLDTMAA 420  
 QY 421 IVNEARIPACSLSSH 435  
 DB 421 IVNEARIPACSLSSH 435

## RESULT 3

AAE22836

ID AAE22836 standard; Protein: 440 AA.

XX AC AAE22836;

XX DT 21-AUG-2002 (first entry)

XX DE Escherichia coli phytase protein.

XX KW dietary aid; biocompatible composition; therapeutic; digestive tract;  
 foodstuff; digestion; phytase; enzyme.  
 XX OS Escherichia coli.

XX DN WO200189317-A2.

XX PP 29-NOV-2001.

XX PR 15-MAY-2001; 2001WO-US15764.

XX PR 25 MAY 2000; 2000US-0580937.

XX PA (DIVER-) DIVERSA CORP.

XX PI Short JM, Kretz KA, O'Donoghue E.

XX DR WPI: 2002-054149/21.

XX DR N-PSDB: AAD36473.

XX PT New dietary aids comprising sustained release biocompatible  
 PT compositions, comprise agent that assists in digestion, useful for  
 PT delivering enzymes, therapeutics, medicine or agents to an organism  
 XX PS Claim 5; Fig 1; 89pp; English.

XX CC The present invention relates to novel dietary aids comprising sustained  
 CC release biocompatible composition which comprises an agent (enzymes such  
 CC as phytase, amylase, esterase, protease) that assists in digestion. The  
 CC biocompatible composition is effective upon oral consumption and release  
 CC in the digestive tract of a subject. The dietary aids are useful for  
 CC delivering enzymes, therapeutics, medicine and agents to an organism.  
 CC The use of enzymes and other agents in digestive aids of livestock or  
 CC domesticated animals not only improves the animal's health and life  
 CC expectancy but also assists in increasing the health of livestock or  
 CC in the production of foodstuffs from livestock. The present sequence  
 CC is Escherichia coli phytase protein.

XX SQ Sequence 440 AA.

Query Match 96.0%; Score 2188; DB 23; Length 440;  
 Best Local Similarity 97.7%; Pred. Neg. 1.2E-212;  
 Matches 425; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

C7: MEALIPFLSLPGLPQSAFAACSEPELKSWSVVSHECVAFATKATLQMDVTHDAWE 60  
 DB 1 MEALIPFLSLPGLPQSAFAACSEPELKSWSVVSHECVAFATKATLQMDVTHDAWE 60

QY 61 TWVVKLGLTPPGSEL IAYLGHYQQTIVADGLAKKSGPQSQVAL IADVIFRTKTGE 120  
 DB 61 TWVVKLGLTPPGSEL IAYLGHYQQTIVADGLAKKSGPQSQVAL IADVIFRTKTGE 120  
 QY 121 APAAGLADPCATIVHTQADTSSSDPDLFNPIKTVGCOLNANVTDAILERAGASTIAFTSH 180  
 DB 121 APAAGLADPCATIVHTQADTSSSDPDLFNPIKTVGCOLNANVTDAILERAGASTIAFTSH 180  
 QY 181 VGLAFSELEPVNPPQSNLTETPEKMEESNSTGALPSTFKVSACVASEETAVPTASMLT 240  
 DB 181 VGLAFSELEPVNPPQSNLTETPEKMEESNSTGALPSTFKVSACVASEETAVPTASMLT 240  
 QY 241 ETEFLQQAQGMPEGVVPTSVIETAGHDTNLANIGGALFNWTLPGGPTNTPGGSEIVFERW 300  
 DB 241 ETEFLQQAQGMPEGVVPTSVIETAGHDTNLANIGGALFNWTLPGGPTNTPGGSEIVFERW 300  
 QY 301 LTPHPPOKQAYGVTLPTSVIETAGHDTNLANIGGALFNWTLPGGPTNTPGGSEIVFERW 360  
 DB 301 LTPHPPOKQAYGVTLPTSVIETAGHDTNLANIGGALFNWTLPGGPTNTPGGSEIVFERW 360  
 QY 361 PPSNSQWIAVSVIETGSHQWSTETETIRACFYLCTETPEVAFSEATETLDTMAA 420  
 DB 361 PPSNSQWIAVSVIETGSHQWSTETETIRACFYLCTETPEVAFSEATETLDTMAA 420  
 QY 421 IVNEARIPACSLSSH 435  
 DB 421 IVNEARIPACSLSSH 435

## RESULT 4

AAE15806

ID AAE15806 standard; Protein: 440 AA.

XX AC AAE15806;

XX DT 26-MAR-2002 (first entry)

XX DE Escherichia coli B phytase protein.

XX KW Bacterial phytase, K12 appA phytase, phytase stability, metabolic;  
 KW gastrointestinal, nutritional value, feed treatment process, therapy;  
 KW thermal tolerance; growth performance, alcoholic drink; biopulping;  
 KW non-alcoholic drink; biobleaching, B phytase.

XX OS Escherichia coli

XX PN WO200190333-A2.

XX PP 29-NOV-2001.

XX PR 24-MAY-2001; 2001WO-US17118.

XX PR 25-MAY-2000; 2000US-0580515

XX PA (DIVER-) DIVERSA CORP.

XX PI Short JM, Kretz KA, Gray KA, Earlen RF, Jarrett JB, O'Donoghue E;

XX DR WPI: 2002-054149/21

XX DR N-PSDB: AAD25460

XX PT New bacterial phytase for e.g. improving the nutritional value of  
 PT phytate-containing foodstuffs and subsequently improving the growth  
 PT performance of an organism that consumes it, or in treating animal  
 PT digestive systems

XX PS Claim 1; Fig 1; 170pp; English.

XX CC The patent discloses recombinant bacterial phytase from Escherichia coli  
 CC K12 appA phytase. The enzyme has phytase activity, and improved thermal  
 CC tolerance when compared with wild type phytase. It has improved protease  
 CC stability at low pH. The recombinant phytase is useful for improving the

cc nutritional value of phytate containing foodstuffs and subsequently  
cc improving the growth performance of an organism that consumes it, in  
cc treating animal digestive systems, in feed treatment processes and for  
cc in vitro purposes related to research, discovery and development; they  
cc are also used for generating recombinant digestive system life forms,  
cc for producing or manufacturing alcoholic and non-alcoholic drinks based  
cc on the use of moulds, grains and/or plants, in biopulping and bio-  
cc bleaching where a reduction in the use of environmentally harmful  
cc chemicals that are traditionally used in the pulp and paper industry  
cc is desired and in the reduction or possible elimination of the need  
cc for mineral supplements, enzymes or therapeutic drugs for animals  
cc from the daily feed thus increasing the amount calories and nutrients  
cc present in the feed. The present sequence is E. coli B phytase protein.

XX Sequence 440 AA;

Query Match 96.08; Score 2188; DB 23; Length 440;  
Best Local Similarity 97.7%; Prod. No. 1.2c-212;  
Matches 425; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1 MKAILIFPLSLIPITQSAFAQSEPEIKLESVVIVSPHGVAPATKATQLMQVTPDAMP 60  
DB 1 MKAILIFPLSLIPITQSAFAQSEPEIKLESVVIVSPHGVAPATKATQLMQVTPDAMP 60  
QY 61 TWVVKLGPIIPRGELIAYLGHYWRQPLVAAGLLPKGQCPQSSQVAIIAIVDERTRKTGE 120  
DB 61 TWVVKLGPIIPRGELIAYLGHYWRQPLVAAGLLPKGQCPQSSQVAIIAIVDERTRKTGE 120  
QY 121 AFAGLAGAPKATIVHTQADTSSPPDLPNPKLTGWCQIDNANVTDAIIERAGGSLADFTGH 180  
DB 121 AFAGLAGAPKATIVHTQADTSSPPDLPNPKLTGWCQIDNANVTDAIIERAGGSLADFTGH 180  
QY 181 YQIAFRPEFVNPQSNICIKPKQDPSQSIQAIDPSEIKVSAIYVSEIYAVSLASMLT 240  
DB 181 YQIAFRPEFVNPQSNICIKPKQDPSQSIQAIDPSEIKVSAIYVSEIYAVSLASMLT 240  
QY 241 EPIELQAQCMPEPGWGRITTSQWNTLSLHNAQFILLQETPEVARSRAPIILLDLKTA 300  
DB 241 EPIELQAQCMPEPGWGRITTSQWNTLSLHNAQFILLQETPEVARSRAPIILLDLKTA 300  
QY 301 LTPHPUKQAYGVTLPTSVLFIAGHDINLANGLALNWLTPGQDPTPPGGHIVFERW 360  
DB 301 LTPHPUKQAYGVTLPTSVLFIAGHDINLANGLALNWLTPGQDPTPPGGHIVFERW 360  
QY 461 RRLSDNSQWTVQSVLFTQDMPEKTLPSLNTPEPEVKLTLAGSEFNASQMSLAGFTQ 420  
DB 461 RRLSDNSQWTVQSVLFTQDMPEKTLPSLNTPEPEVKLTLAGSEFNASQMSLAGFTQ 420  
QY 421 LVNEARIPAGSLKSH 445  
DB 421 LVNEARIPAGSLKSH 445

RESULT 5  
AAH77775  
ID AAH77775 standard; Protein: 442 AA.

XX AAH77775;

DI 05-JUN-2002 (first entry)

XX Phytase associated protein.

XX Phytase.

XX Unidentified.

XX KR99086028-A.

XX 15-DEC-1999.

XX 25-MAY-1998; 98KR-0018810.

XX

PR 25-MAY-1998; 98KR-0018810.

XX (W30J-1) W30JIN CO LTD.

XX Bae HD, Forecobarth CW, Goloban S, Chent KJ;

XX WPI; 2000-045078/62.

XX N-PSDB; ARK12514.

XX Novel phytase gene, recombinant phytase and usage thereof

XX Disclosure; Fig 3; lopp; Korean.

XX The invention relates to a novel phytase gene, a recombinant

XX phytase gene and their uses. This is the amino acid sequence of the

XX phytase associated protein described in the invention.

XX Sequence 442 AA;

Query Match 95.7%; Score 2182; DB 21; Length 442;

Best Local Similarity 98.1%; Prod. No. 4.9c-272;

Matches 424; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 MKAILIFPLSLIPITQSAFAQSEPEIKLESVVIVSPHGVAPATKATQLMQVTPDAMP 60

DB 1 MKAILIFPLSLIPITQSAFAQSEPEIKLESVVIVSPHGVAPATKATQLMQVTPDAMP 60

QY 61 TWVVKLGPIIPRGELIAYLGHYWRQPLVAAGLLPKGQCPQSSQVAIIAIVDERTRKTGE 120

DB 61 TWVVKLGPIIPRGELIAYLGHYWRQPLVAAGLLPKGQCPQSSQVAIIAIVDERTRKTGE 120

QY 121 AFAGLAGAPKATIVHTQADTSSPPDLPNPKLTGWCQIDNANVTDAIIERAGGSLADFTGH 180

DB 121 AFAGLAGAPKATIVHTQADTSSPPDLPNPKLTGWCQIDNANVTDAIIERAGGSLADFTGH 180

QY 181 YQIAFRPEFVNPQSNICIKPKQDPSQSIQAIDPSEIKVSAIYVSEIYAVSLASMLT 240

DB 181 YQIAFRPEFVNPQSNICIKPKQDPSQSIQAIDPSEIKVSAIYVSEIYAVSLASMLT 240

QY 241 EPIELQAQCMPEPGWGRITTSQWNTLSLHNAQFILLQETPEVARSRAPIILLDLKTA 300

DB 241 EPIELQAQCMPEPGWGRITTSQWNTLSLHNAQFILLQETPEVARSRAPIILLDLKTA 300

QY 301 LTPHPUKQAYGVTLPTSVLFIAGHDINLANGLALNWLTPGQDPTPPGGHIVFERW 360

DB 301 LTPHPUKQAYGVTLPTSVLFIAGHDINLANGLALNWLTPGQDPTPPGGHIVFERW 360

QY 461 RRLSDNSQWTVQSVLFTQDMPEKTLPSLNTPEPEVKLTLAGSEFNASQMSLAGFTQ 420

DB 461 RRLSDNSQWTVQSVLFTQDMPEKTLPSLNTPEPEVKLTLAGSEFNASQMSLAGFTQ 420

QY 421 LVNEARIPAGSLKSH 442

DB 421 LVNEARIPAGSLKSH 442

RESULT 6

AAH46257

ID AAB46257 standard; Protein: 442 AA.

XX AAB46257;

XX 20-FEB-2001 (first entry)

XX Lama2/Abpa plasmid translated sequences.

XX Transgenic animal; salivary protein; phytase; phosphatase animal, growth

XX environmental pollution; pig.

XX Mus musculus - chimeric.

XX Escherichia coli - chimeric.

XX WC200064247-A1.



```

XX 02-NOV-2000.
PD
XX
XX 20-APR-2000; 2000WO-CA00430.
PF
XX
XX 24-APR-1999; 99US-0130508.
PR
XX
XX (UYGU) UNIV GUELPH.
PA
XX
XX Forsberg CW, Golovan S, Phillips JP;
PI
XX
XX WPI; 2000-687245/67.
DR
XX
XX N-PSDB; AAC68294.
DR
XX
XX Transgenic non-human animal for gastrointestinal tract specific
PT expression of a protein, preferably phytase, comprises a nucleic acid
PT sequence including a heterologous transgene construct encoding the
PT protein.
XX
XX Dislosure; Fig 5; 152pp; English.
PS
XX
XX The present invention provides transgenic animals which produce desired
CC proteins, in this case pigs which expresses phytase in the salivary
CC gland. Low phytase production levels result in phytate in the diet being
CC excreted and causing phosphorus contamination in water, as well as
CC reducing the growth of animals. The invention provides a number of
CC transgenes containing the E. coli APPA phytase coding sequence.
XX
XX Sequence 432 AA:
Query Match 95.7%; Score 2182; DB 21; Length 432.
Best Local Similarity 98.1%; Pred. No. 4.80-212;
Matches 424; Conservative 0; Mismatches 8; Indels 0; Gaps 0.
QY 1 MKAIIIPFLLSLIPITPQSAFAQSPPELKLSVIVSRHGVAPTRATQIMQDVIPDAMP 60
DB 1 MKAIIIPFLLSLIPITPQSAFAQSPPELKLSVIVSRHGVAPTRATQIMQDVIPDAMP 60
QY 61 TWPVKLGWLTTPGGRI IAVIGHYQKQPIVAQGI IAKKGTQGSQVAIIAIWVEPEPKTGE 120
DB 61 TWPVKLGWLTTPGGRI IAVIGHYQKQPIVAQGI IAKKGTQGSQVAIIAIWVEPEPKTGE 120
QY 121 AFAAGLAPGCAIVHTQADTSPPDIPNPKTGVQOLDNANVTDAILERAGGSIAPFTGH 180
DB 121 AFAAGLAPGCAIVHTQADTSPPDIPNPKTGVQOLDNANVTDAILERAGGSIAPFTGH 180
QY 181 YOTAFPELFRVLFNFTQSNLCTEPPEKDESCSLTQALPSELKVSADNVSITGAVSLASMLT 240
DB 181 YOTAFPELFRVLFNFTQSNLCTEPPEKDESCSLTQALPSELKVSADNVSITGAVSLASMLT 240
QY 241 EFLAQAGAMPPEQWGRITTSQWNTIISLHNAQFYLLQPTPEVAPSPATPLDLIKTA 300
DB 241 EFLAQAGAMPPEQWGRITTSQWNTIISLHNAQFYLLQPTPEVAPSPATPLDLIKTA 300
QY 301 LTPHPQKQAYGVTLPTSVLFTAGHDNLANLGAGLELNWTLPGQPDNTPPGGELVFERW 360
DB 301 LTPHPQKQAYGVTLPTSVLFTAGHDNLANLGAGLELNWTLPGQPDNTPPGGELVFERW 360
QY 361 PRLSUNSWLQVSLVFCILQCMPEKPIKLSINTPPEVKTIACCEPERNACMCSLAGFTQ 420
DB 361 PRLSUNSWLQVSLVFCILQCMPEKPIKLSINTPPEVKTIACCEPERNACMCSLAGFTQ 420

```

```

DT 20-FEB-2001 (first entry)
XX
XX R15/APPA plasmid translated sequence.
DE
XX
XX Transgenic animal; salivary protein; phytase; phosphorus; animal growth;
KW environmental pollution; pig.
XX
XX Rattus sp - chimeric.
OS
XX Escherichia coli - chimeric.
XX
XX WO2000064247-A1.
PN
XX
XX 02-NOV-2000.
PD
XX
XX 20-APR-2000; 2000WO-CA00430.
PF
XX
XX 24-APR-1999; 99US-0130508.
PR
XX
XX (UYGU) UNIV GUELPH.
PA
XX
XX Forsberg CW, Golovan S, Phillips JP;
PI
XX
XX WPI; 2000-687245/67.
DR
XX
XX N-PSDB; AAC68295.
DR
XX
XX Transgenic non-human animal for gastrointestinal tract specific
PT expression of a protein, preferably phytase, comprises a nucleic acid
PT sequence including a heterologous transgene construct encoding the
PT protein.
XX
XX Dislosure; Fig 18; 152pp; English.
PS
XX
XX The present invention provides transgenic animals which produce desired
CC proteins, in this case pigs which expresses phytase in the salivary
CC gland. Low phytase production levels result in phytate in the diet being
CC excreted and causing phosphorus contamination in water, as well as
CC reducing the growth of animals. The invention provides a number of
CC transgenes containing the E. coli APPA phytase coding sequence.
XX
XX Sequence 432 AA:
Query Match 95.7%; Score 2182; DB 21; Length 432.
Best Local Similarity 98.1%; Pred. No. 4.80-212;
Matches 424; Conservative 0; Mismatches 8; Indels 0; Gaps 0.
QY 1 MKAIIIPFLLSLIPITPQSAFAQSPPELKLSVIVSRHGVAPTRATQIMQDVIPDAMP 60
DB 1 MKAIIIPFLLSLIPITPQSAFAQSPPELKLSVIVSRHGVAPTRATQIMQDVIPDAMP 60
QY 61 TWPVKLGWLTTPGGRI IAVIGHYQKQPIVAQGI IAKKGTQGSQVAIIAIWVEPEPKTGE 120
DB 61 TWPVKLGWLTTPGGRI IAVIGHYQKQPIVAQGI IAKKGTQGSQVAIIAIWVEPEPKTGE 120
QY 121 AFAAGLAPGCAIVHTQADTSPPDIPNPKTGVQOLDNANVTDAILERAGGSIAPFTGH 180
DB 121 AFAAGLAPGCAIVHTQADTSPPDIPNPKTGVQOLDNANVTDAILERAGGSIAPFTGH 180
QY 181 YOTAFPELFRVLFNFTQSNLCTEPPEKDESCSLTQALPSELKVSADNVSITGAVSLASMLT 240
DB 181 YOTAFPELFRVLFNFTQSNLCTEPPEKDESCSLTQALPSELKVSADNVSITGAVSLASMLT 240
QY 241 EFLAQAGAMPPEQWGRITTSQWNTIISLHNAQFYLLQPTPEVAPSPATPLDLIKTA 300
DB 241 EFLAQAGAMPPEQWGRITTSQWNTIISLHNAQFYLLQPTPEVAPSPATPLDLIKTA 300
QY 301 LTPHPQKQAYGVTLPTSVLFTAGHDNLANLGAGLELNWTLPGQPDNTPPGGELVFERW 360
DB 301 LTPHPQKQAYGVTLPTSVLFTAGHDNLANLGAGLELNWTLPGQPDNTPPGGELVFERW 360
QY 361 PRLSUNSWLQVSLVFCILQCMPEKPIKLSINTPPEVKTIACCEPERNACMCSLAGFTQ 420
DB 361 PRLSUNSWLQVSLVFCILQCMPEKPIKLSINTPPEVKTIACCEPERNACMCSLAGFTQ 420

```

```

QY 421 IVNEARIPACSL 432
DB 421 IVNEARIPACSL 432

RESULT 8
AAB36259
ID AAB36259 standard; Protein: 432 AA.
AC AAB36259;
XX
DT 20-FEB-2001 (first entry)
DE R15/APPA plasmid translated sequence
KW Transgenic animal; salivary protein; phytase; phosphorus; animal growth;
KW environmental pollution; pig.
OS Rattus sp - chimeric.
XX Escherichia coli - chimeric.
XX
DN W0200064247-A1.
XX
PD 02-NOV-2000.
XX
PF 20-APR-2000; 2000WO-CA00430.
XX
PR 23-APR-1999; 990S-0140508.
XX
PA (UYGU-) UNIV GUELPH.
XX
PI Forsberg CW, Golovan S, Phillips JP;
XX
DR WPI: 2000-687245/67.
XX N PSDB: AAC68296.
XX
PT Transgenic non-human animal for gastrointestinal tract specific
PT expression of a protein, preferably phytase, comprises a nucleic acid
PT sequence including a heterologous transgene construct encoding the
PT protein
XX
PS Disclosure: Fig 19; 152pp; English.
XX
CC The present invention provides transgenic animals which produce desired
CC proteins, in this case pigs which express phytase in the salivary
CC gland. Low phytase production levels result in phytate in the diet being
CC excreted and causing phosphorus contamination in water, as well as
CC reducing the growth of animals. The invention provides a number of
CC transgenes containing the E. coli APPA phytase coding sequence.
XX
SQ Sequence 432 AA;

Query Match 66.7%; Score 2182; DB 21; Length 432;
Best Local Similarity 98.1%; Pred. No. 4,86-212;
Matches 424; Conservative 0; Mismatches 8; Indels 0; Gaps 0.

QY 1 MKAILPFLSLILPTPSAFAQSEPELKLESVVIVSHGHVAPTKATQLMQDVTTPDAMP 60
DB 1 MKAILPFLSLILPTPSAFAQSEPELKLESVVIVSHGHVAPTKATQLMQDVTTPDAMP 60
QY 61 TWPVKLGELTPRGGFIAYIGYVWPGQVAVAGCLIPKGTGPGSGGVAIIAPVDEPTKGTGE 120
DB 61 TWPVKLGELTPRGGFIAYIGYVWPGQVAVAGCLIPKGTGPGSGGVAIIAPVDEPTKGTGE 120
QY 121 AFAPGIFAPVAVITVHTADPTSSPPPLFNPIKTVVQVLNANVTDAITFPAQGSSTADPTGH 180
DB 121 AFAGGLAFDCAITVHTQADTSSPPPLFNPIKTVGQVLNANVTDAITLSPAGSSSTADPTGH 180
QY 181 YQTAFPRFVNLNPPQSNLCIKPKQDESLSITQALPSKLVSDCVSLTQAVSLASMLT 240
DB 181 KQTAFPRFVNLNPPQSNLCIKRKQDESLSITQALPSKLVSDCNVSLTGAVSLASMLT 240
QY 241 FIFLIQAQGMPEPCWGRITDTSQWNTLILSHNAQFYLLQRTPEVARSKATPLILDLKTA 300

```

```

DB 241 FIFLIQAQGMPEPCWGRITDTSQWNTLILSHNAQFYLLQRTPEVARSKATPLILDLKTA 300
QY 301 LTPHPQKQAYGVTLPTSVLFTAGHDTNLANAGGALIEHNLWLLFGSDPTTFPSSEAVFEKW 400
DB 301 LTPHPQKQAYGVTLPTSVLFTAGHDTNLANAGGALIEHNLWLLFGSDPTTFPSSEAVFEKW 400
QY 361 KKLSSNSQWLVQSVLVLQGLQGMKDKPLSSINPLPGGEVALLLAGLLEKSNAGSMSSAGPTG 420
DB 361 KKLSSNSQWLVQSVLVLQGLQGMKDKPLSSINPLPGGEVALLLAGLLEKSNAGSMSSAGPTG 420
QY 421 IVNEARIPACSL 432
DB 421 IVNEARIPACSL 432

RESULT 9
AAB36260
ID AAB36260 standard; Protein: 432 AA.
XX
AC AAB36260;
XX
DT 20-FEB-2001 (first entry)
DE R15/APPA plasmid translated sequence.
KW Transgenic animal; salivary protein; phytase; phosphorus; animal growth;
KW environmental pollution; pig.
OS Rattus sp - chimeric.
XX Escherichia coli - chimeric.
XX
DN W0200064247-A1.
XX
PD 02-NOV-2000.
XX
PF 20-APR-2000; 2000WO-CA00430.
XX
PR 23-APR-1999; 990S-0140508.
XX
PA (UYGU-) UNIV GUELPH.
XX
PI Forsberg CW, Golovan S, Phillips JP;
XX
DR WPI: 2000-687245/67.
XX N PSDB: AAC68297.
XX
PT Transgenic non-human animal for gastrointestinal tract specific
PT expression of a protein, preferably phytase, comprises a nucleic acid
PT sequence including a heterologous transgene construct encoding the
PT protein
XX
PS Disclosure: Fig 20; 152pp; English.
XX
CC The present invention provides transgenic animals which produce desired
CC proteins, in this case pigs which express phytase in the salivary
CC gland. Low phytase production levels result in phytate in the diet being
CC excreted and causing phosphorus contamination in water, as well as
CC reducing the growth of animals. The invention provides a number of
CC transgenes containing the E. coli APPA phytase coding sequence.
XX
SQ Sequence 432 AA;

Query Match 95.7%; Score 2182; DB 21; Length 432;
Best Local Similarity 98.1%; Pred. No. 4,86-212;
Matches 424; Conservative 0; Mismatches 8; Indels 0; Gaps 0.

QY 1 MKAILPFLSLILPTPSAFAQSEPELKLESVVIVSHGHVAPTKATQLMQDVTTPDAMP 60
DB 1 MKAILPFLSLILPTPSAFAQSEPELKLESVVIVSHGHVAPTKATQLMQDVTTPDAMP 60
QY 61 TWPVKLGELTPRGGFIAYIGYVWPGQVAVAGCLIPKGTGPGSGGVAIIAPVDEPTKGTGE 120

```

DB 61 TWPKVLEMLTTPGDEH IAYI GHYQOPRI VAISEII IAKKQIPQSUGVAIIADVDERTKTGE 120  
QY 121 AFAAGLAPDCAITVHTQADTSSPDPLFNPLKTVGCOLDNANVTDAILLERAGGSTADFTGH 180  
DB 121 AFAAGLAPDCAITVHTQADTSSPDPLFNPLKTVGCOLDNANVTDAILLERAGGSTADFTGH 180  
QY 181 YGTAFRELEPRVINFPSQNLCKPPEQPS' S'ITCAIPSELKVSADNVSITCAVSLASMLT 240  
DB 181 YGTAFRELEPRVINFPSQNLCKPPEQPS' S'ITCAIPSELKVSADNVSITCAVSLASMLT 240  
QY 241 EIEFLQQAQGMPEPCWGRITTSHOWNTLISHNAQFYLQRTPEVARSRAITPLDILKTA 300  
DB 241 EIEFLQQAQGMPEPCWGRITTSHOWNTLISHNAQFYLQRTPEVARSRAITPLDILKTA 300  
QY 301 LTPHPPOKQAYGVTLPTSVLFTAGHDNTNLANLGGALELNWTLPGQPDNTPPGGELVFERW 360  
DB 301 LTPHPPOKQAYGVTLPTSVLFTAGHDNTNLANLGGALELNWTLPGQPDNTPPGGELVFERW 360  
QY 361 FRLSDNSQWTOVSLVFQTLQGMFKTPI SINTPPGEVKI TLACGEFERNACQMSIAGFTQ 420  
DB 361 FRLSDNSQWTOVSLVFQTLQGMFKTPI SINTPPGEVKI TLACGEFERNACQMSIAGFTQ 420  
QY 421 IVNEARIPACSL 432  
DB 421 IVNEARIPACSL 432  
RESULT 10  
AAB36261  
ID AAB36261 standard; Protein: 432 AA.  
AC AAB36261;  
UT 20-FEB-2001 (first entry)  
XX RL5/APPA plasmid translated sequence.  
DE Transgenic animal; salivary protein; phytase; phosphorus; animal growth;  
KW environmental pollution; pig.  
KW Rattus sp - chimeric.  
OS Escherichia coli - chimeric.  
XX Wo2000064247-A1.  
XX 02-NOV-2000.  
XX 20-APR-2000; 2000WO-CA00430  
XX 23-APR-1999; 99US-0130508  
XX (UYGU-) UNIV GUELPH.  
XX Forsberg CW, Golovan S, Phillips JP;  
XX WPI: 2000-687245/67.  
XX N-PSDB: AAC58298.  
XX Transgenic non-human animal for gastrointestinal tract specific  
PT expression of a protein, preferably phytase, comprises a nucleic acid  
PT sequence including a heterologous transgene construct encoding the  
PT protein -  
XX Disclosure: Fig 21; 152pp; English.  
XX The present invention provides transgenic animals which produce desired  
XX proteins, in this case pigs which expresses phytase in the salivary  
XX gland, low phytase production levels result in phytate in the diet being  
XX excreted and causing phosphorus contamination in water, as well as  
XX reducing the growth of animals. The invention provides a number of  
XX transgenes containing the E. coli APPA phytase coding sequence.  
XX Sequence 432 AA;  
SQ

Query Match 95.7%; Score 2182; Db 21; length 432;  
Best Local Similarity 98.1%; Pred. No. 4,8e-212;  
Matches 424; Conservative 0; Mismatches 8; Indels 0; Gaps 0;  
QY 1 MKAILIPFLSILLPLTPQSAFAQSEPKLIESVIVSRHCVRAPTKATQIMQIVHTDAMP 60  
DB 1 MKAILIPFLSILLPLTPQSAFAQSEPKLIESVIVSRHCVRAPTKATQIMQIVHTDAMP 60  
QY 61 TWPKVLEMLTTPGDEH IAYI GHYQOPRI VAISEII IAKKQIPQSUGVAIIADVDERTKTGE 120  
DB 61 TWPKVLEMLTTPGDEH IAYI GHYQOPRI VAISEII IAKKQIPQSUGVAIIADVDERTKTGE 120  
QY 121 AFAAGLAPDCAITVHTQADTSSPDPLFNPLKTVGCOLDNANVTDAILLERAGGSTADFTGH 180  
DB 121 AFAAGLAPDCAITVHTQADTSSPDPLFNPLKTVGCOLDNANVTDAILLERAGGSTADFTGH 180  
QY 181 YGTAFRELEPRVINFPSQNLCKPPEQPS' S'ITCAIPSELKVSADNVSITCAVSLASMLT 240  
DB 181 YGTAFRELEPRVINFPSQNLCKPPEQPS' S'ITCAIPSELKVSADNVSITCAVSLASMLT 240  
QY 241 EIEFLQQAQGMPEPCWGRITTSHOWNTLISHNAQFYLQRTPEVARSRAITPLDILKTA 300  
DB 241 EIEFLQQAQGMPEPCWGRITTSHOWNTLISHNAQFYLQRTPEVARSRAITPLDILKTA 300  
QY 301 LTPHPPOKQAYGVTLPTSVLFTAGHDNTNLANLGGALELNWTLPGQPDNTPPGGELVFERW 360  
DB 301 LTPHPPOKQAYGVTLPTSVLFTAGHDNTNLANLGGALELNWTLPGQPDNTPPGGELVFERW 360  
QY 361 FRLSDNSQWTOVSLVFQTLQGMFKTPI SINTPPGEVKI TLACGEFERNACQMSIAGFTQ 420  
DB 361 FRLSDNSQWTOVSLVFQTLQGMFKTPI SINTPPGEVKI TLACGEFERNACQMSIAGFTQ 420  
QY 421 IVNEARIPACSL 432  
DB 421 IVNEARIPACSL 432  
RESULT 11  
AAB36262  
ID AAB36262 standard; Protein: 432 AA.  
AC AAB36262;  
UT 20-FEB-2001 (first entry)  
XX SV40/APPA plasmid translated sequence.  
DE Transgenic animal; salivary protein; phytase; phosphorus; animal growth;  
KW environmental pollution; pig.  
XX Rhesus macaque polyoma virus - chimeric.  
OS Escherichia coli - chimeric.  
XX Wo2000064247-A1.  
XX 02-NOV-2000.  
XX 20-APR-2000; 2000WO-CA00430.  
XX 23-APR-1999; 99US-0130508.  
XX (UYGU-) UNIV GUELPH.  
XX Forsberg CW, Golovan S, Phillips JP;  
XX WPI: 2000-687245/67.  
XX N-PSDB: AAC68299.  
XX Transgenic non-human animal for gastrointestinal tract specific  
PT expression of a protein, preferably phytase, comprises a nucleic acid  
PT sequence including a heterologous transgene construct encoding the  
PT protein -



non-alcoholic drink; biobatching.

Escherichia coli.

W0200190333-A2.

29-NOV-2001.

24-MAY-2001: 2001W0-US17118.

25-MAY-2000: 2000US-0540515.

(DIVE-) DIVERSA CORP.

Short JM, Kretz KA, Gray KA, Barton NP, Garrett JH, O'Donoghue E;

WPI: 2002-083108/11.

N-PSDB; AAD25463.

Now bacterial phytase for e.g. improving the nutritional value of phytate-containing foodstuffs and subsequently improving the growth performance of an organism that consumes it, or in treating animal digestive systems

claim 54; Fig 8; 170pp; English.

The patent discloses recombinant bacterial phytase from Escherichia coli E.244A phytase. The enzyme has phytase activity and improved thermal tolerance when compared with wild type phytase. It has improved protease stability at low pH. The recombinant phytase is useful for improving the nutritional value of phytate-containing foodstuffs and subsequently improving the growth performance of an organism that consumes it, in treating animal digestive systems, in feed treatment processes and for in vitro purposes related to research, discovery and development. They are also used for generating recombinant digestive system like forms, for producing or manufacturing alcoholic and non-alcoholic drinks based on the use of moulds, grains and/or plants, in biopulping and bio-bleaching where a reduction in the use of environmentally harmful chemicals that are traditionally used in the pulp and paper industry is desired and in the reduction or possible elimination of the need for mineral supplements, enzymes or therapeutic drugs for animals from the daily feed thus increasing the amount calories and nutrients present in the feed. The present sequence is E. coli appA phytase wild type protein.

Query Match 95.7%, Score 2192; DB 23; Length 432.

Best Local Similarity 98.1%; Pred. No. 4, 8c-212;

Matches 424; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 MKALLPFLSLILPLTPQSAFAQSEPELKLESVVVSRHGVRAPTRKATQIMQVILDAWF 60

DB 1 MKALLPFLSLILPLTPQSAFAQSEPELKLESVVVSRHGVRAPTRKATQIMQVILDAWF 60

QY 61 TWPVKLGWLPKGLTAYLGHYQORLVAQGLLAKKQCPQSGVALLIADVDETRKTGE 120

DB 61 TWPVKLGWLPKGLTAYLGHYQORLVAQGLLAKKQCPQSGVALLIADVDETRKTGE 120

QY 121 AFANANAPVATVHTAGVSCRPENPIKTVYQVQIDNANVATLTPAGGSGIADPTGH 180

DB 121 AFANANAPVATVHTAGVSCRPENPIKTVYQVQIDNANVATLTPAGGSGIADPTGH 180

QY 181 YCTAPPELPVLPNPSNICTKPKQDFSCITQATPSEIKVSDCVSLTCAVSLASMLT 240

DB 181 YCTAPPELPVLPNPSNICTKPKQDFSCITQATPSEIKVSDCVSLTCAVSLASMLT 240

QY 241 EFLTQQAGMPPEPCWGPITUSHOWNTLILSHNAQFDLLOPTPKVARSATPLDLTKTA 300

DB 241 EFLTQQAGMPPEPCWGPITUSHOWNTLILSHNAQFDLLOPTPKVARSATPLDLTKTA 300

QY 401 LTPHPPOKQAVGVTLSVLFIAQHDNTLANLGGALFLNWTLPQDPNTPGGSLVFERW 360

DB 401 LTPHPPOKQAVGVTLSVLFIAQHDNTLANLGGALFLNWTLPQDPNTPGGSLVFERW 360

301 LTPHPPOKQAVGVTLSVLFIAQHDNTLANLGGALFLNWTLPQDPNTPGGSLVFERW 360

361 RRLSDNSQWITQVSLVFTLQOMKDKTFLSLNTPPEVKLFIAQHDNTLANLGGALFLNWTLPQDPNTPGGSLVFERW 420

361 RRLSDNSQWITQVSLVFTLQOMKDKTFLSLNTPPEVKLFIAQHDNTLANLGGALFLNWTLPQDPNTPGGSLVFERW 420

421 IVNEARIPACSL 432

421 IVNEARIPACSL 432

RESULT 14

AAE02631

ID AAE02631 standard, Protein; 432 AA.

XX AC AAE02631;

XX DT 06-AUG-2001 (first entry)

XX DE E. coli acid phosphatase/phytase (appA) protein.

XX AC acid phosphatase-phytase; appA, enzymatic activity; food additive;

KW animal feed; monoester phosphate; cereal food; human food.

XX OS Escherichia coli.

XX FH Key location/Qualifiers

FT Misc-difference 109 /note= "Encoded by GPC"

FT Misc-difference 148 /note= "Encoded by ATT"

FT Modified-site 161..163 /note= "Asn is N-glycosylated"

FT Disulfide-bond 200..210

FT Region 202..211 /label= GH\_loop

FT Modified-site 339..341 /note= "Asn is N-glycosylated"

XX W0200136607-A1

XX 25-MAY-2001.

XX 17-NOV-2000: 2000W0-US31622.

XX 18-NOV-1999: 99US-0166179.

XX (COPP ) CORNELL RES FOUND INC.

XX Lei X;

WPI: 2001-367572/48

N-PSDB; AAD06831.

Mutated acid phosphatase/phytase from Escherichia coli has improved enzymatic activity compared to the wild type and is useful as a food additive, particularly for animal feeds

Claim 1; Page 11-12; 56pp; English.

The present sequence is wild type acid phosphatase/phytase (appA) from Escherichia coli. The wild type appA is mutated to enhance its enzymatic activity. Phytases, a specific group of monoester phosphatases, are required to initiate the release of phosphate from phosphates, the major storage of phosphate in cereal foods or feeds. The mutant acid phosphatase/phytase is added to animal feed as a food additive to improve uptake of phosphate and zinc from the diet. The phytase can also be added to human food. The mutant phytase is more heat stable than phytase isolated from Aspergillus niger and safer for use in human food manufacture

Sequence 432 AA;

Query Match 95.59; Score 2177; DP 22; Length 432;  
 Best Local Similarity 97.78; Pred. No. 6,26,211;  
 Matches 422; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 MKATLIPFSLSLIPITPQSAFAQSFPFLKLESVVIVSRHCVAPTRATQLMQVTPWMP 60  
 DQ 1 MKATLIPFSLSLIPITPQSAFAQSFPFLKLESVVIVSRHCVAPTRATQLMQVTPWMP 60

QY 61 TWVVKIGETTPRGFTIAYIGRWQDRI VADGHI PKQSGTQSCQVALIAVWDEPTKTF 120  
 DQ 61 TWVVKIGETTPRGFTIAYIGRWQDRI VADGHI PKQSGTQSCQVALIAVWDEPTKTF 120

QY 121 AFAGLAPKATIVHTQATSSPFTFENLKTGWQDENANVTIATIERAGSSIADEIGH 180  
 DQ 121 AFAGLAPKATIVHTQATSSPFTFENLKTGWQDENANVTIATIERAGSSIADEIGH 180

QY 181 YQAFRELEEVLPQSNICLKREKQDESSELTOALPSELKVSADCVSLIGAVSLASMLT 240  
 DQ 181 YQAFRELEEVLPQSNICLKREKQDESSELTOALPSELKVSADCVSLIGAVSLASMLT 240

QY 241 ETELQQAQCMPEPGWGRITDSHOWNTLSIHNAQFDLQPTDEVAPSRATPLILDKTA 300  
 DQ 241 ETELQQAQCMPEPGWGRITDSHOWNTLSIHNAQFDLQPTDEVAPSRATPLILDKTA 300

QY 301 LIDPPQKQAYGVITPVSLEFLAHCNTANLGGTALFIRWITPQENIPFGSELVFERW 360  
 DQ 301 LIDPPQKQAYGVITPVSLEFLAHCNTANLGGTALFIRWITPQENIPFGSELVFERW 360

QY 361 RRLSDNSQWLVQSVLFTQCMPTKRTLSINTPPGVKRTILAGSEFENAGQCMFSLAGFTQ 420  
 DQ 361 RRLSDNSQWLVQSVLFTQCMPTKRTLSINTPPGVKRTILAGSEFENAGQCMFSLAGFTQ 420

QY 421 TVNEARIPACSL 432  
 DQ 421 TVNEARIPACSL 432

RESULT 15  
 AAE02644  
 ID AAE02644 standard; Protein: 432 AA.  
 AC AAE02644;  
 DT 06-AUG-2001 (first entry)  
 DE E. coli acid phosphatase/phytase (appa) mutant (A131N).  
 KW Acid phosphatase-phytase; appa; enzymatic activity; mutant; mutagen;  
 KW animal food; monoster phosphate; coral food; human food; food additive  
 XX Escherichia coli.  
 OS Synthetic.  
 XX Key Location/Qualifiers  
 FT Misc-difference 14;  
 FT /note- "Wild type Ala substituted with Asn"

Search completed: December 14, 2002, 15:53:43  
 Job time : 42 secs

PT additive, particularly for animal foods  
 XX Example 1: Page 5; 5pp; English.  
 XX The present sequence is Escherichia coli acid phosphatase/phytase (appa) mutant (A131N). The wild type appa is mutated to enhance its enzymatic activity. Phytases, a specific group of monoester phosphates, are required to initiate the release of phosphate from phytate, the major storage of phosphate in cereal foods or feeds. The mutant acid phosphatase/phytase is added to animal food as a food additive to improve uptake of phosphate and zinc from the diet. The phytase can also be added to human food. The mutant phytase is more heat stable than phytase isolated from Aspergillus niger and safer for use in human food manufacture.  
 CC Note: the present sequence is not shown in the specification but is derived from wild appa referred as SEQ ID NO: 2 (AA02644) and shown on page 6, 7 of the specification.  
 XX Sequence 432 AA;  
 SS

Query Match 95.59; Score 2177; DP 22; Length 432;  
 Best Local Similarity 97.78; Pred. No. 6,26,211;  
 Matches 422; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 MKATLIPFSLSLIPITPQSAFAQSFPFLKLESVVIVSRHCVAPTRATQLMQVTPWMP 60  
 DQ 1 MKATLIPFSLSLIPITPQSAFAQSFPFLKLESVVIVSRHCVAPTRATQLMQVTPWMP 60

QY 61 TWVVKIGETTPRGFTIAYIGRWQDRI VADGHI PKQSGTQSCQVALIAVWDEPTKTF 120  
 DQ 61 TWVVKIGETTPRGFTIAYIGRWQDRI VADGHI PKQSGTQSCQVALIAVWDEPTKTF 120

QY 121 AFAGLAPKATIVHTQATSSPFTFENLKTGWQDENANVTIATIERAGSSIADEIGH 180  
 DQ 121 AFAGLAPKATIVHTQATSSPFTFENLKTGWQDENANVTIATIERAGSSIADEIGH 180

QY 181 YQAFRELEEVLPQSNICLKREKQDESSELTOALPSELKVSADCVSLIGAVSLASMLT 240  
 DQ 181 YQAFRELEEVLPQSNICLKREKQDESSELTOALPSELKVSADCVSLIGAVSLASMLT 240

QY 241 ETELQQAQCMPEPGWGRITDSHOWNTLSIHNAQFDLQPTDEVAPSRATPLILDKTA 300  
 DQ 241 ETELQQAQCMPEPGWGRITDSHOWNTLSIHNAQFDLQPTDEVAPSRATPLILDKTA 300

QY 301 LIDPPQKQAYGVITPVSLEFLAHCNTANLGGTALFIRWITPQENIPFGSELVFERW 360  
 DQ 301 LIDPPQKQAYGVITPVSLEFLAHCNTANLGGTALFIRWITPQENIPFGSELVFERW 360

QY 361 RRLSDNSQWLVQSVLFTQCMPTKRTLSINTPPGVKRTILAGSEFENAGQCMFSLAGFTQ 420  
 DQ 361 RRLSDNSQWLVQSVLFTQCMPTKRTLSINTPPGVKRTILAGSEFENAGQCMFSLAGFTQ 420

QY 421 TVNEARIPACSL 432  
 DQ 421 TVNEARIPACSL 432

GenCore version 5.1.3  
Copyright (c) 1993 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 13, 2002, 15:49:55 : Search time 13 Seconds  
(without alignments)  
1391.052 Million cell updates/sec

Title: US-09-866-379a-10

Perfect score: 2279

Sequence: 1 MKAILPFLSLPLTPQSA : GFTQIVNEAFIPACSLPSHL 436

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum db seq length: 200000000

Post processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2182	95.7	422	PPA_ECOLI	P07103 escherichia
2	554	24.3	413	AGF_ECOLI	P19926 escherichia
3	552	24.2	413	AGP_SALTY	O33921 salmonella
4	538	23.6	417	AGP_PROPE	Q52309 providencia
5	134.5	5.9	423	PPAL_HUMAN	P11117 homo sapien
6	119	5.2	423	PPAL_RAT	P20611 rattus norv
7	111.5	4.9	421	PPAL_MOUSE	P24638 mus musculu
8	111.5	4.9	457	PAX8_MOUSE	Q00288 mus musculu
9	109.5	4.8	479	PHYB_ASFAP	P34755 aspergillus
10	108.5	4.8	479	PHYB_ASFAP	P34754 aspergillus
11	108	4.7	386	PPAP_HUMAN	P15309 homo sapien
12	107.5	4.7	459	PAX8_CANFA	P47240 canis famil
13	105	4.6	450	PAX8_HUMAN	Q06710 homo sapien
14	103.5	4.5	1048	SBCC_ECOLI	P13458 escherichia
15	103	4.5	769	SVBL_CACCE	Q04116 escherichia
16	101.5	4.5	635	TPOR_HUMAN	P40238 homo sapien
17	103	4.4	371	PPAP_RAT	P20645 rattus norv
18	98.5	4.3	381	PPAP_RAT	P20645 rattus norv
19	96.5	4.2	328	PXAL_DSCAL	Q01504 pseudomonas
20	95	4.2	2534	TLEL_PORPE	P13368 drosophila
21	94	4.1	341	TIN2_MOUSE	Q04899 mus musculu
22	94	4.1	612	MYEF_XFPA	Q01573 zebrafish
23	94	4.1	755	PPAX_CACCE	Q04549 caenorhabdi
24	94	4.1	895	DAG1_BOVIN	O18738 bos taurus
25	94.5	4.1	972	OCG4_SCHPO	Q07794 schistosom
26	93	4.1	625	TPOR_MOUSE	Q08451 mus musculu
27	92.5	4.1	763	PPAR_HUMAN	Q08460 homo sapien
28	92.5	4.1	1085	CASR_BOVIN	P35384 bos taurus
29	91.5	4.0	291	PPC_PFYCO	P16101 corythibac
30	91.5	4.0	872	PPAL_RAT	P42335 rattus norv
31	91	4.0	458	PAX8_RAT	P51971 rattus norv
32	91	4.0	551	CAP_MOUSE	P36521 schizosacch
33	91	4.0	638	SRPR_HUMAN	P08240 homo sapien

34 91 4.0 811 1 PPA\_RHORI  
35 91 4.0 1374 1 VCAP\_HSV11  
36 91 4.0 3033 1 POLG\_HCV66  
37 90.5 4.0 369 1 HEM3\_PEA  
38 90.5 4.0 467 1 PHYA\_ASFAP  
39 90.5 4.0 1464 1 NCO2\_HUMAN  
40 90 3.9 457 1 CD4\_MOUSE  
41 90 3.9 461 1 ATOC\_ECOLI  
42 90 3.9 668 1 YFCK\_ECOLI  
43 90 3.9 921 1 TOS3\_DROME  
44 90 3.9 1079 1 CASP\_MOUSE  
45 89.5 3.9 394 1 PPAY\_CAEEL

P05445 rhodospiril  
P06491 herpes simp  
P26650 h genome po  
Q4082 pisum sativ  
P34754 aspergillus  
Q15596 homo sapien  
P06432 mus musculu  
Q06065 escherichia  
P77182 escherichia  
P17825 drosophila  
Q94996 mus musculu  
Q10944 caenorhabdi

## ALIGNMENTS

RESULT 1  
PPA\_ECOLI STANDARD: PRT: 432 AA.  
AC P07102  
DT 01-APR-1988 (rel. 07, Created)  
DT 01-AUG-1991 (rel. 19, Last sequence update)  
DE periplasmic appa protein precursor [includes: Phosphatohydride  
DE phosphohydrolase (EC 3.1.3.26)] (PH 2.5 acid phosphatase) (AP); 6-  
DE phytase (EC 3.1.3.26)]  
CN APPA OR P0840.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria, gamma subdivision, Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM H.A., AND SEQUENCE OF 23-29.  
RC STRAIN-K12;  
EX MEDLINE=90368616; PubMed=2168385;  
RA Dassa J., Marek C., Boquet P.L.;  
RT "The complete nucleotide sequence of the Escherichia coli gene appa  
RT reveals significant homology between pH 2.5 acid phosphatase and  
RT glucose-1-phosphatase."  
RL J. Bacteriol. 172:5497-5509(1990).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-K12 / MG1655;  
EX MEDLINE=97426617; PubMed=9278503;  
RA Blatner F.R., Plunkett G., III, Beach C.A., Fouts N.T., Burland V.,  
RA Riley M., Collado-Vides J., Glasner J.P., Rode C.K., Mayhew G.P.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
RA Mau B., Shao Y.;  
RT "The complete genome sequence of Escherichia coli K-12."  
RL Science 277:1232-1243(1997).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-K12;  
EX MEDLINE=97662202; PubMed=8905232;  
RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,  
RA Ikemoto K., Itoh T., Katoh N., Kubota N., Makino K., Mochizuki K.,  
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Miyajima K.,  
RA Mori H., Motomizu K., Nakamura Y., Nakamura Y., Saito R.,  
RA Campi C., Seki Y., Tanami H., Takemata K., Wada C., Yamamoto Y.,  
RA Yano M., Horiiuchi T.;  
RT "A 7.9-kb DNA sequence of the Escherichia coli K-12 genome  
RT contains information for the 12729 genes, including the appa gene."  
RL DNA Res. 3:137-155(1996).  
RN [4]  
RP SEQUENCE OF 1-112 FROM N.A.  
EX MEDLINE=8721756; PubMed=3038201;  
RA Tonati E., Danchin A.;  
RT "The structure of the promoter and amino terminal region of the pH  
RT 2.5 acid phosphatase structural gene (appa) of E. coli: a negative  
RT control of transcription mediated by cyclic AMP."  
RL Biochimie 69:215-221(1987).













```

DB 8 QAAALQFLGCMGLTVMPTQARS-----LPFTLLPHGDRSPVKT----YKDP 53
QY 57 DAMPTWVKLGELTPGGELIATLGHWRORIVADGLPKCGGPGSQGVAILADVD-ERT 115
DB 54 YQEKWPGQGLTKGSMLOHWEGLQALFQY--HGF-----NTSYHPQFVVYSTDEPT 108
QY 116 RKTGEFAAGLADCAITVHTOADTSSPDLFNLPLKTCVQIQDANVTAILEPAGGSIA 175
DB 109 LMSAEANLAGIFPP-----NEVQHESPNISWQPI-----PVHTVITE----- 146
QY 176 DFTGHVQTAFRELEVRVINEPQSNLCIKPEKQDESCSLTQALPSLKLKACVSLTCAVSL 235
DB 147 -----DRLLKPELPGCPREYQONE-----TROTPEYQNPISQNAQFLNMVAN 186
QY 236 ASMLTEHPG-----LOGAQMPEPGWG-----PITUSHOWN-----TLLSLNAQF 276
DB 190 ETGLTNVLTETINWVITLFCETHGLLLPPWASPTVQPLSLKURSELFLFGLHE--- 246
QY 277 DLORTFEVARS-----ATPLDLIKTALTTPHPQKQAVGVILDTSVLFIAGHJTNL 329
DB 247 -----QVKARLQGGVILAQIKNLIIMATTSQFPK-----LLVYSAHDITL 288
QY 330 ANLCMALEINLWTLFQPTNTPFCTVFEFEPFESFNSGWIQVSLVPLQGLQMFELFSL 387
DB 289 VALQMALNVY-----NCKQAFYASCHIFELYQ EDNGNF SVDMYFRDSK --KAPWP 337
QY 390 LNTTPGVEKVLTLAGCHERRAQQMCISLAGTQIVNERAP 428
DB 338 LILP-----GCPHR-----CPLQDFLRL-TERVIP 361

RESULT 8
PAX8 MOUSE
ID PAX8 MOUSE STANDARD: PRT: 457 AA.
AC Q00248; P97342;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Paired box protein Pax-8.
GN PAX8 OR PAX-8.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN 1
RP SEQUENCE FROM N.A.
RX MEDLINE-92155118; PubMed 1723970;
RA Plachow D., Chowdhury K., Walther C., Simon D., Guenet J.-L.,
RA Gross P.;
RT "pax8, a murine paired box gene expressed in the developing excretory
RT system and thyroid gland.";
RL development 110:643-651(1990).
RN 121
RP REVIEWS TO 158; 263 AND 340.
RX MEDLINE-92387536; PubMed 1516825;
PA Adams B., Boerker P., Agazzi A., Kmitz Z., Urbank P.,
PA Maurer-Poy F., Busslinger M.;
RT "pax-5 encodes the transcription factor ISAP and is expressed in B
RT lymphocytes, the developing CNS, and adult testis.";
RL Genes Dev. 6:1589-1607(1992).
RN 131
RP SEQUENCE FROM N.A.
RX STRAIN C57BL/6; TISSUE-Kidney;
RX MEDLINE-97349113; PubMed-9205117;
PA Okladnova O., Poloc A., Pantes J., Lee M., Plachow P., Horst T.,
PA "The genomic organization of the murine Pax 8 gene and
RT characterization of its basal promoter.";
RL Genomics 42:452-461(1997).
CC -!- FUNCTION: THOUGHT TO ENCODE A TRANSCRIPTION FACTOR. IT MAY HAVE A
CC ROLE IN KIDNEY CELL DIFFERENTIATION. MAY PLAY A REGULATORY ROLE IN
CC MAMMALIAN DEVELOPMENT.
CC -!- SUBCELLULAR LOCATION: Nuclear.

```

```

CC -!- TISSUE SPECIFICITY: EXPRESSED IN THE DEVELOPING EXCRETORY
CC SYSTEM AND THE THYROID GLAND.
CC -!- SIMILARITY: CONTAINS 1 PAIRED BOX DOMAIN.
CC
CC THIS SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed, used by and for commercial
CC entities to pursue a license, other than for the purpose of making a copy
CC or send an email to licensee@sib.ch.
CC
EMBL: X57487; CAA40725.1; ALT_SEQ.
EMBL: X95592; CAA67904.1;
EMBL: X95593; CAA67904.1; JOINED.
EMBL: X95594; CAA67904.1; JOINED.
EMBL: X95595; CAA67904.1; JOINED.
EMBL: X95596; CAA67904.1; JOINED.
EMBL: X95597; CAA67904.1; JOINED.
EMBL: X95598; CAA67904.1; JOINED.
EMBL: X95599; CAA67904.1; JOINED.
EMBL: X95600; CAA67904.1; JOINED.
EMBL: X95601; CAA67904.1; JOINED.
EMBL: X95602; CAA67904.1; JOINED.
EMBL: X95603; CAA67904.1; JOINED.
EMBL: X95604; CAA67904.1; JOINED.
EMBL: X95605; CAA67904.1; JOINED.
EMBL: X95606; CAA67904.1; JOINED.
EMBL: X95607; CAA67904.1; JOINED.
EMBL: X95608; CAA67904.1; JOINED.
EMBL: X95609; CAA67904.1; JOINED.
EMBL: X95610; CAA67904.1; JOINED.
EMBL: X95611; CAA67904.1; JOINED.
EMBL: X95612; CAA67904.1; JOINED.
EMBL: X95613; CAA67904.1; JOINED.
EMBL: X95614; CAA67904.1; JOINED.
EMBL: X95615; CAA67904.1; JOINED.
EMBL: X95616; CAA67904.1; JOINED.
EMBL: X95617; CAA67904.1; JOINED.
EMBL: X95618; CAA67904.1; JOINED.
EMBL: X95619; CAA67904.1; JOINED.
EMBL: X95620; CAA67904.1; JOINED.
EMBL: X95621; CAA67904.1; JOINED.
EMBL: X95622; CAA67904.1; JOINED.
EMBL: X95623; CAA67904.1; JOINED.
EMBL: X95624; CAA67904.1; JOINED.
EMBL: X95625; CAA67904.1; JOINED.
EMBL: X95626; CAA67904.1; JOINED.
EMBL: X95627; CAA67904.1; JOINED.
EMBL: X95628; CAA67904.1; JOINED.
EMBL: X95629; CAA67904.1; JOINED.
EMBL: X95630; CAA67904.1; JOINED.
EMBL: X95631; CAA67904.1; JOINED.
EMBL: X95632; CAA67904.1; JOINED.
EMBL: X95633; CAA67904.1; JOINED.
EMBL: X95634; CAA67904.1; JOINED.
EMBL: X95635; CAA67904.1; JOINED.
EMBL: X95636; CAA67904.1; JOINED.
EMBL: X95637; CAA67904.1; JOINED.
EMBL: X95638; CAA67904.1; JOINED.
EMBL: X95639; CAA67904.1; JOINED.
EMBL: X95640; CAA67904.1; JOINED.
EMBL: X95641; CAA67904.1; JOINED.
EMBL: X95642; CAA67904.1; JOINED.
EMBL: X95643; CAA67904.1; JOINED.
EMBL: X95644; CAA67904.1; JOINED.
EMBL: X95645; CAA67904.1; JOINED.
EMBL: X95646; CAA67904.1; JOINED.
EMBL: X95647; CAA67904.1; JOINED.
EMBL: X95648; CAA67904.1; JOINED.
EMBL: X95649; CAA67904.1; JOINED.
EMBL: X95650; CAA67904.1; JOINED.
EMBL: X95651; CAA67904.1; JOINED.
EMBL: X95652; CAA67904.1; JOINED.
EMBL: X95653; CAA67904.1; JOINED.
EMBL: X95654; CAA67904.1; JOINED.
EMBL: X95655; CAA67904.1; JOINED.
EMBL: X95656; CAA67904.1; JOINED.
EMBL: X95657; CAA67904.1; JOINED.
EMBL: X95658; CAA67904.1; JOINED.
EMBL: X95659; CAA67904.1; JOINED.
EMBL: X95660; CAA67904.1; JOINED.
EMBL: X95661; CAA67904.1; JOINED.
EMBL: X95662; CAA67904.1; JOINED.
EMBL: X95663; CAA67904.1; JOINED.
EMBL: X95664; CAA67904.1; JOINED.
EMBL: X95665; CAA67904.1; JOINED.
EMBL: X95666; CAA67904.1; JOINED.
EMBL: X95667; CAA67904.1; JOINED.
EMBL: X95668; CAA67904.1; JOINED.
EMBL: X95669; CAA67904.1; JOINED.
EMBL: X95670; CAA67904.1; JOINED.
EMBL: X95671; CAA67904.1; JOINED.
EMBL: X95672; CAA67904.1; JOINED.
EMBL: X95673; CAA67904.1; JOINED.
EMBL: X95674; CAA67904.1; JOINED.
EMBL: X95675; CAA67904.1; JOINED.
EMBL: X95676; CAA67904.1; JOINED.
EMBL: X95677; CAA67904.1; JOINED.
EMBL: X95678; CAA67904.1; JOINED.
EMBL: X95679; CAA67904.1; JOINED.
EMBL: X95680; CAA67904.1; JOINED.
EMBL: X95681; CAA67904.1; JOINED.
EMBL: X95682; CAA67904.1; JOINED.
EMBL: X95683; CAA67904.1; JOINED.
EMBL: X95684; CAA67904.1; JOINED.
EMBL: X95685; CAA67904.1; JOINED.
EMBL: X95686; CAA67904.1; JOINED.
EMBL: X95687; CAA67904.1; JOINED.
EMBL: X95688; CAA67904.1; JOINED.
EMBL: X95689; CAA67904.1; JOINED.
EMBL: X95690; CAA67904.1; JOINED.
EMBL: X95691; CAA67904.1; JOINED.
EMBL: X95692; CAA67904.1; JOINED.
EMBL: X95693; CAA67904.1; JOINED.
EMBL: X95694; CAA67904.1; JOINED.
EMBL: X95695; CAA67904.1; JOINED.
EMBL: X95696; CAA67904.1; JOINED.
EMBL: X95697; CAA67904.1; JOINED.
EMBL: X95698; CAA67904.1; JOINED.
EMBL: X95699; CAA67904.1; JOINED.
EMBL: X95700; CAA67904.1; JOINED.
EMBL: X95701; CAA67904.1; JOINED.
EMBL: X95702; CAA67904.1; JOINED.
EMBL: X95703; CAA67904.1; JOINED.
EMBL: X95704; CAA67904.1; JOINED.
EMBL: X95705; CAA67904.1; JOINED.
EMBL: X95706; CAA67904.1; JOINED.
EMBL: X95707; CAA67904.1; JOINED.
EMBL: X95708; CAA67904.1; JOINED.
EMBL: X95709; CAA67904.1; JOINED.
EMBL: X95710; CAA67904.1; JOINED.
EMBL: X95711; CAA67904.1; JOINED.
EMBL: X95712; CAA67904.1; JOINED.
EMBL: X95713; CAA67904.1; JOINED.
EMBL: X95714; CAA67904.1; JOINED.
EMBL: X95715; CAA67904.1; JOINED.
EMBL: X95716; CAA67904.1; JOINED.
EMBL: X95717; CAA67904.1; JOINED.
EMBL: X95718; CAA67904.1; JOINED.
EMBL: X95719; CAA67904.1; JOINED.
EMBL: X95720; CAA67904.1; JOINED.
EMBL: X95721; CAA67904.1; JOINED.
EMBL: X95722; CAA67904.1; JOINED.
EMBL: X95723; CAA67904.1; JOINED.
EMBL: X95724; CAA67904.1; JOINED.
EMBL: X95725; CAA67904.1; JOINED.
EMBL: X95726; CAA67904.1; JOINED.
EMBL: X95727; CAA67904.1; JOINED.
EMBL: X95728; CAA67904.1; JOINED.
EMBL: X95729; CAA67904.1; JOINED.
EMBL: X95730; CAA67904.1; JOINED.
EMBL: X95731; CAA67904.1; JOINED.
EMBL: X95732; CAA67904.1; JOINED.
EMBL: X95733; CAA67904.1; JOINED.
EMBL: X95734; CAA67904.1; JOINED.
EMBL: X95735; CAA67904.1; JOINED.
EMBL: X95736; CAA67904.1; JOINED.
EMBL: X95737; CAA67904.1; JOINED.
EMBL: X95738; CAA67904.1; JOINED.
EMBL: X95739; CAA67904.1; JOINED.
EMBL: X95740; CAA67904.1; JOINED.
EMBL: X95741; CAA67904.1; JOINED.
EMBL: X95742; CAA67904.1; JOINED.
EMBL: X95743; CAA67904.1; JOINED.
EMBL: X95744; CAA67904.1; JOINED.
EMBL: X95745; CAA67904.1; JOINED.
EMBL: X95746; CAA67904.1; JOINED.
EMBL: X95747; CAA67904.1; JOINED.
EMBL: X95748; CAA67904.1; JOINED.
EMBL: X95749; CAA67904.1; JOINED.
EMBL: X95750; CAA67904.1; JOINED.
EMBL: X95751; CAA67904.1; JOINED.
EMBL: X95752; CAA67904.1; JOINED.
EMBL: X95753; CAA67904.1; JOINED.
EMBL: X95754; CAA67904.1; JOINED.
EMBL: X95755; CAA67904.1; JOINED.
EMBL: X95756; CAA67904.1; JOINED.
EMBL: X95757; CAA67904.1; JOINED.
EMBL: X95758; CAA67904.1; JOINED.
EMBL: X95759; CAA67904.1; JOINED.
EMBL: X95760; CAA67904.1; JOINED.
EMBL: X95761; CAA67904.1; JOINED.
EMBL: X95762; CAA67904.1; JOINED.
EMBL: X95763; CAA67904.1; JOINED.
EMBL: X95764; CAA67904.1; JOINED.
EMBL: X95765; CAA67904.1; JOINED.
EMBL: X95766; CAA67904.1; JOINED.
EMBL: X95767; CAA67904.1; JOINED.
EMBL: X95768; CAA67904.1; JOINED.
EMBL: X95769; CAA67904.1; JOINED.
EMBL: X95770; CAA67904.1; JOINED.
EMBL: X95771; CAA67904.1; JOINED.
EMBL: X95772; CAA67904.1; JOINED.
EMBL: X95773; CAA67904.1; JOINED.
EMBL: X95774; CAA67904.1; JOINED.
EMBL: X95775; CAA67904.1; JOINED.
EMBL: X95776; CAA67904.1; JOINED.
EMBL: X95777; CAA67904.1; JOINED.
EMBL: X95778; CAA67904.1; JOINED.
EMBL: X95779; CAA67904.1; JOINED.
EMBL: X95780; CAA67904.1; JOINED.
EMBL: X95781; CAA67904.1; JOINED.
EMBL: X95782; CAA67904.1; JOINED.
EMBL: X95783; CAA67904.1; JOINED.
EMBL: X95784; CAA67904.1; JOINED.
EMBL: X95785; CAA67904.1; JOINED.
EMBL: X95786; CAA67904.1; JOINED.
EMBL: X95787; CAA67904.1; JOINED.
EMBL: X95788; CAA67904.1; JOINED.
EMBL: X95789; CAA67904.1; JOINED.
EMBL: X95790; CAA67904.1; JOINED.
EMBL: X95791; CAA67904.1; JOINED.
EMBL: X95792; CAA67904.1; JOINED.
EMBL: X95793; CAA67904.1; JOINED.
EMBL: X95794; CAA67904.1; JOINED.
EMBL: X95795; CAA67904.1; JOINED.
EMBL: X95796; CAA67904.1; JOINED.
EMBL: X95797; CAA67904.1; JOINED.
EMBL: X95798; CAA67904.1; JOINED.
EMBL: X95799; CAA67904.1; JOINED.
EMBL: X95800; CAA67904.1; JOINED.
EMBL: X95801; CAA67904.1; JOINED.
EMBL: X95802; CAA67904.1; JOINED.
EMBL: X95803; CAA67904.1; JOINED.
EMBL: X95804; CAA67904.1; JOINED.
EMBL: X95805; CAA67904.1; JOINED.
EMBL: X95806; CAA67904.1; JOINED.
EMBL: X95807; CAA67904.1; JOINED.
EMBL: X95808; CAA67904.1; JOINED.
EMBL: X95809; CAA67904.1; JOINED.
EMBL: X95810; CAA67904.1; JOINED.
EMBL: X95811; CAA67904.1; JOINED.
EMBL: X95812; CAA67904.1; JOINED.
EMBL: X95813; CAA67904.1; JOINED.
EMBL: X95814; CAA67904.1; JOINED.
EMBL: X95815; CAA67904.1; JOINED.
EMBL: X95816; CAA67904.1; JOINED.
EMBL: X95817; CAA67904.1; JOINED.
EMBL: X95818; CAA67904.1; JOINED.
EMBL: X95819; CAA67904.1; JOINED.
EMBL: X95820; CAA67904.1; JOINED.
EMBL: X95821; CAA67904.1; JOINED.
EMBL: X95822; CAA67904.1; JOINED.
EMBL: X95823; CAA67904.1; JOINED.
EMBL: X95824; CAA67904.1; JOINED.
EMBL: X95825; CAA67904.1; JOINED.
EMBL: X95826; CAA67904.1; JOINED.
EMBL: X95827; CAA67904.1; JOINED.
EMBL: X95828; CAA67904.1; JOINED.
EMBL: X95829; CAA67904.1; JOINED.
EMBL: X95830; CAA67904.1; JOINED.
EMBL: X95831; CAA67904.1; JOINED.
EMBL: X95832; CAA67904.1; JOINED.
EMBL: X95833; CAA67904.1; JOINED.
EMBL: X95834; CAA67904.1; JOINED.
EMBL: X95835; CAA67904.1; JOINED.
EMBL: X95836; CAA67904.1; JOINED.
EMBL: X95837; CAA67904.1; JOINED.
EMBL: X95838; CAA67904.1; JOINED.
EMBL: X95839; CAA67904.1; JOINED.
EMBL: X95840; CAA67904.1; JOINED.
EMBL: X95841; CAA67904.1; JOINED.
EMBL: X95842; CAA67904.1; JOINED.
EMBL: X95843; CAA67904.1; JOINED.
EMBL: X95844; CAA67904.1; JOINED.
EMBL: X95845; CAA67904.1; JOINED.
EMBL: X95846; CAA67904.1; JOINED.
EMBL: X95847; CAA67904.1; JOINED.
EMBL: X95848; CAA67904.1; JOINED.
EMBL: X95849; CAA67904.1; JOINED.
EMBL: X95850; CAA67904.1; JOINED.
EMBL: X95851; CAA67904.1; JOINED.
EMBL: X95852; CAA67904.1; JOINED.
EMBL: X95853; CAA67904.1; JOINED.
EMBL: X95854; CAA67904.1; JOINED.
EMBL: X95855; CAA67904.1; JOINED.
EMBL: X95856; CAA67904.1; JOINED.
EMBL: X95857; CAA67904.1; JOINED.
EMBL: X95858; CAA67904.1; JOINED.
EMBL: X95859; CAA67904.1; JOINED.
EMBL: X95860; CAA67904.1; JOINED.
EMBL: X95861; CAA67904.1; JOINED.
EMBL: X95862; CAA67904.1; JOINED.
EMBL: X95863; CAA67904.1; JOINED.
EMBL: X95864; CAA67904.1; JOINED.
EMBL: X95865; CAA67904.1; JOINED.
EMBL: X95866; CAA67904.1; JOINED.
EMBL: X95867; CAA67904.1; JOINED.
EMBL: X95868; CAA67904.1; JOINED.
EMBL: X95869; CAA67904.1; JOINED.
EMBL: X95870; CAA67904.1; JOINED.
EMBL: X95871; CAA67904.1; JOINED.
EMBL: X95872; CAA67904.1; JOINED.
EMBL: X95873; CAA67904.1; JOINED.
EMBL: X95874; CAA67904.1; JOINED.
EMBL: X95875; CAA67904.1; JOINED.
EMBL: X95876; CAA67904.1; JOINED.
EMBL: X95877; CAA67904.1; JOINED.
EMBL: X95878; CAA67904.1; JOINED.
EMBL: X95879; CAA67904.1; JOINED.
EMBL: X95880; CAA67904.1; JOINED.
EMBL: X95881; CAA67904.1; JOINED.
EMBL: X95882; CAA67904.1; JOINED.
EMBL: X95883; CAA67904.1; JOINED.
EMBL: X95884; CAA67904.1; JOINED.
EMBL: X95885; CAA67904.1; JOINED.
EMBL: X95886; CAA67904.1; JOINED.
EMBL: X95887; CAA67904.1; JOINED.
EMBL: X95888; CAA67904.1; JOINED.
EMBL: X95889; CAA67904.1; JOINED.
EMBL: X95890; CAA67904.1; JOINED.
EMBL: X95891; CAA67904.1; JOINED.
EMBL: X95892; CAA67904.1; JOINED.
EMBL: X95893; CAA67904.1; JOINED.
EMBL: X95894; CAA67904.1; JOINED.
EMBL: X95895; CAA67904.1; JOINED.
EMBL: X95896; CAA67904.1; JOINED.
EMBL: X95897; CAA67904.1; JOINED.
EMBL: X95898; CAA67904.1; JOINED.
EMBL: X95899; CAA67904.1; JOINED.
EMBL: X95900; CAA67904.1; JOINED.
EMBL: X95901; CAA67904.1; JOINED.
EMBL: X95902; CAA67904.1; JOINED.
EMBL: X95903; CAA67904.1; JOINED.
EMBL: X95904; CAA67904.1; JOINED.
EMBL: X95905; CAA67904.1; JOINED.
EMBL: X95906; CAA67904.1; JOINED.
EMBL: X95907; CAA67904.1; JOINED.
EMBL: X95908; CAA67904.1; JOINED.
EMBL: X95909; CAA67904.1; JOINED.
EMBL: X95910; CAA67904.1; JOINED.
EMBL: X95911; CAA67904.1; JOINED.
EMBL: X95912; CAA67904.1; JOINED.
EMBL: X95913; CAA67904.1; JOINED.
EMBL: X95914; CAA67904.1; JOINED.
EMBL: X95915; CAA67904.1; JOINED.
EMBL: X95916; CAA67904.1; JOINED.
EMBL: X95917; CAA67904.1; JOINED.
EMBL: X95918; CAA67904.1; JOINED.
EMBL: X95919; CAA67904.1; JOINED.
EMBL: X95920; CAA67904.1; JOINED.
EMBL: X95921; CAA67904.1; JOINED.
EMBL: X95922; CAA67904.1; JOINED.
EMBL: X95923; CAA67904.1; JOINED.
EMBL: X95924; CAA67904.1; JOINED.
EMBL: X95925; CAA67904.1; JOINED.
EMBL: X95926; CAA67904.1; JOINED.
EMBL: X95927; CAA67904.1; JOINED.
EMBL: X95928; CAA67904.1; JOINED.
EMBL: X95929; CAA67904.1; JOINED.
EMBL: X95930; CAA67904.1; JOINED.
EMBL: X95931; CAA67904.1; JOINED.
EMBL: X95932; CAA67904.1; JOINED.
EMBL: X95933; CAA67904.1; JOINED.
EMBL: X95934; CAA67904.1; JOINED.
EMBL: X95935; CAA67904.1; JOINED.
EMBL: X95936; CAA67904.1; JOINED.
EMBL: X95937; CAA67904.1; JOINED.
EMBL: X95938; CAA67904.1; JOINED.
EMBL: X95939; CAA67904.1; JOINED.
EMBL: X95940; CAA67904.1; JOINED.
EMBL: X95941; CAA67904.1; JOINED.
EMBL: X95942; CAA67904.1; JOINED.
EMBL: X95943; CAA67904.1; JOINED.
EMBL: X95944; CAA67904.1; JOINED.
EMBL: X95945; CAA67904.1; JOINED.
EMBL: X95946; CAA67904.1; JOINED.
EMBL: X95947; CAA67904.1; JOINED.
EMBL: X95948; CAA67904.1; JOINED.
EMBL: X95949; CAA67904.1; JOINED.
EMBL: X95950; CAA67904.1; JOINED.
EMBL: X95951; CAA67904.1; JOINED.
EMBL: X95952; CAA67904.1; JOINED.
EMBL: X95953; CAA67904.1; JOINED.
EMBL: X95954; CAA67904.1; JOINED.
EMBL: X95955; CAA67904.1; JOINED.
EMBL: X95956; CAA67904.1; JOINED.
EMBL: X95957; CAA67904.1; JOINED.
EMBL: X95958; CAA67904.1; JOINED.
EMBL: X95959; CAA67904.1; JOINED.
EMBL: X95960; CAA67904.1; JOINED.
EMBL: X95961; CAA67904.1; JOINED.
EMBL: X95962; CAA67904.1; JOINED.
EMBL: X95963; CAA67904.1; JOINED.
EMBL: X95964; CAA67904.1; JOINED.
EMBL: X95965; CAA67904.1; JOINED.
EMBL: X95966; CAA67904.1; JOINED.
EMBL: X95967; CAA67904.1; JOINED.
EMBL: X95968; CAA67904.1; JOINED.
EMBL: X95969; CAA67904.1; JOINED.
EMBL: X95970; CAA67904.1; JOINED.
EMBL: X95971; CAA67904.1; JOINED.
EMBL: X95972; CAA67904.1; JOINED.
EMBL: X95973; CAA67904.1; JOINED.
EMBL: X95974; CAA67904.1; JOINED.
EMBL: X95975; CAA67904.1; JOINED.
EMBL: X95976; CAA67904.1; JOINED.
EMBL: X95977; CAA67904.1; JOINED.
EMBL: X95978; CAA67904.1; JOINED.
EMBL: X95979; CAA67904.1; JOINED.
EMBL: X95980; CAA67904.1; JOINED.
EMBL: X95981; CAA67904.1; JOINED.
EMBL: X95982; CAA67904.1; JOINED.
EMBL: X95983; CAA67904.1; JOINED.
EMBL: X95984; CAA67904.1; JOINED.
EMBL: X95985; CAA67904.1; JOINED.
EMBL: X95986; CAA67904.1; JOINED.
EMBL: X95987; CAA67904.1; JOINED.
EMBL: X95988; CAA67904.1; JOINED.
EMBL: X95989; CAA67904.1; JOINED.
EMBL: X95990; CAA67904.1; JOINED.
EMBL: X95991; CAA67904.1; JOINED.
EMBL: X95992; CAA67904.1; JOINED.
EMBL: X95993; CAA67904.1; JOINED.
EMBL: X95994; CAA67904.1; JOINED.
EMBL: X95995; CAA67904.1; JOINED.
EMBL: X95996; CAA67904.1; JOINED.
EMBL: X95997; CAA67904.1; JOINED.
EMBL: X95998; CAA67904.1; JOINED.
EMBL: X95999; CAA67904.1; JOINED.
EMBL: X96000; CAA67904.1; JOINED.
EMBL: X96001; CAA67904.1; JOINED.
EMBL: X96002; CAA67904.1; JOINED.
EMBL: X96003; CAA67904.1; JOINED.
EMBL: X96004; CAA67904.1; JOINED.
EMBL: X96005; CAA67904.1; JOINED.
EMBL: X96006; CAA67904.1; JOINED.
EMBL: X96007; CAA67904.1; JOINED.
EMBL: X96008; CAA67904.1; JOINED.
EMBL: X96009; CAA67904.1; JOINED.
EMBL: X96010; CAA67904.1; JOINED.
EMBL: X96011; CAA67904.1; JOINED.
EMBL: X96012; CAA67904.1; JOINED.
EMBL: X96013; CAA67904.1; JOINED.
EMBL: X96014; CAA67904.1; JOINED.
EMBL: X96015; CAA67904.1; JOINED.
EMBL: X96016; CAA67904.1; JOINED.
EMBL: X96017; CAA67904.1; JOINED.
EMBL: X96018; CAA67904.1; JOINED.
EMBL: X96019; CAA67904.1; JOINED.
EMBL: X96020; CAA67904.1; JOINED.
EMBL: X96021; CAA67904.1; JOINED.
EMBL: X96022; CAA67904.1; JOINED.
EMBL: X96023; CAA67904.1; JOINED.
EMBL: X96024; CAA67904.1; JOINED.
EMBL: X96025; CAA67904.1; JOINED.
EMBL: X96026; CAA67904.1; JOINED.
EMBL: X96027; CAA67904.1; JOINED.
EMBL: X96028; CAA67904.1; JOINED.
EMBL: X96029; CAA67904.1; JOINED.
EMBL: X96030; CAA67904.1; JOINED.
EMBL: X96031; CAA67904.1; JOINED.
EMBL: X96032; CAA67904.1; JOINED.
EMBL: X96033; CAA67904.1; JOINED.
EMBL: X96034; CAA67904.1; JOINED.
EMBL: X96035; CAA67904.1; JOINED.
EMBL: X96036; CAA67904.1; JOINED.
EMBL: X96037; CAA67904.1; JOINED.
EMBL: X96038; CAA67904.1; JOINED.
EMBL: X96039; CAA67904.1; JOINED.
EMBL: X96040; CAA67904.1;
```









```

SQ SEQUENCE 459 AA: 48839 MW: 4000164A9E56A935 CR664;
Query Match 4.7%; Score 107.5; DB 1; Length 459;
Best Local Similarity 18.7%; Pred. No. 0.31;
Matches 81; Conservative 51; Mismatches 130; Indels 171; Gaps 16.

QY 26 PELKLESVVIVNSHGVRATKATQI-----MODVTPDAMPVTPVKLCEL-----TP 71
   ||: : : : : |||
DB 27 PEVVRGRIVDLAIDAVPPDITSEQLPVSHGCVSKILCPYETOSIPPTVGGSKPKVATP 80
   ||: : : : : |||
QY 72 REBELIAYLCHWRQ-----RIVANGEL-----LF 95
   ||: : : : : |||
DB 87 K---VVEK(GDYKQKQNPIMFANFIDRLIAAGVCTNDTPVESVSSINFIISTKVOQFENLP 143
   ||: : : : : |||
QY 96 KAC-----POS-----GQVAIADVDEPTPPCEAF 122
   ||: : : : : |||
DB 144 MDSVATKSLSPHTLIPSAVTPPSPQSDISGSTYSINGLIGTAQPSGSDSKPKRMUSD 203
   ||: : : : : |||
QY 123 AAGLAPKCAITVHTQADTSPPDIFNPLKTCVQVQIDANVTVVAILEPACGSIADPTCHVQ 182
   ||: : : : : |||
DB 204 ----QDSQPISTNSQSSSSCP-----KILPTDAESQHLEPLECPPEQ-----HTP 247
   ||: : : : : |||
QY 183 TAFRIERVLNFTQSNLTLEKQKQESLSLQALFSEKVSALCVSL-----TGAV 223
   ||: : : : : |||
DB 248 EAY-----ASPSHTKGEQGLVPLPLNSAALIDGKA 278
   ||: : : : : |||
QY 244 SLASMLTETELIQAGMPPEWCFITDSHWNTLLSLINACFDLLOPTPEVAPSPATP- 292
   ||: : : : : |||
DB 279 TLTPSNTPLE-----GRNSTHTQTVVADVHPSPFAIKQETPEVSSSSSTPS 324
   ||: : : : : |||
QY 293 ----LILIKTALTPHPQKQAVGVTLPTSLVF-----IAGHDINLANLGCALPLNMT 341
   ||: : : : : |||
DB 325 SLSSNAFLDLQVAGVPA-----CASVPPNAPFAASVYQFTQCALISCPENVGPT 379
   ||: : : : : |||
QY 342 LRCQDNTIPDGE 354
   ||: : : : : |||
DB 380 LRCYDPPHPTISQ 392
   ||: : : : : |||

RESULT 14
PAX8_HUMAN
ID PAX8_HUMAN STANDARD: PRT: 450 AA.
AC Q06710; Q09155; Q16337; Q16338; Q16339;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Paired box protein Pax-8.
CN PAX8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN 1;
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), AND VARIANT LEU 329
RC TISSUE=Kidney;
RX MEDLINE=9310167; PubMed=1357742;
RA Folter A., Fickenscher H., Mundlos S., Winterpacht A., Zabel B.,
RA Fidler A., Grass P., Plachov D.,
RT "PAX8, a human paired box gene: Isolation and expression in
RL developing thyroid, kidney and Wilms' tumors.";
RN 121;
RP SEQUENCE FROM N.A. (ISOFORM 1), AND ALTERNATIVE SPLICING.
RC TISSUE=Kidney;
RX MEDLINE=94019277; PubMed 8413205;
RA Kozmik Z., Karchner P., Folter A., Bassilind M.,
RT "Alternative splicing of Pax-8 gene transcripts is developmentally
RT regulated and generates isoforms with different transactivation
RT properties.";
PE Mol. Cell. Biol. 13:6324-6335(1993).
RN 131;
RP SEQUENCE FROM N.A. (ISOFORMS 1; 3; 4 AND 5).
RX MEDLINE=95255300; PubMed 7737192,

```

```

PA Polster A., Weidner P., Fickenscher H., Mundlos M.S., Winterpacht A.,
PA Abbott C., Plachov D.;
RT "Distinct functional properties of three human paired-box proteins,
RT PAX8, isoforms generated by alternative splicing in thyroid, kidney
RT and Wilms' tumors.";
PL Eur. J. Biochem. 228:899-911(1995).
RN 141;
RP VARIANT ID PRO-40;
EX MEDLINE 21394129; PubMed 11562839;
PA Congdon T., Nguyen T.Q., Nequeira C.R., Habbib R.L., Medeiros-Neto G.,
PA Kopp P.;
RT "A novel mutation (Q40P) in PAX8 associated with congenital
RT hypothyroidism and thyroid hypoplasia: evidence for phenotypic
RT variability in mother and child.";
RL J. Clin. Endocrinol. Metab. 90:3562-3567(2001).
RN 151;
RP VARIANTS TO HIS-31 AND ARC-62;
RX MEDLINE=98250178; PubMed 9590296;
PA Mochizuki P.E., Lapi F., Klode H., Fritto M.L., Misero G., Chiovato L.,
PA Soudani A., Baserga M., Tassi V., Pinchera A., Fenzi G., Gruters A.,
PA Busslinger M., Di Lauro R.;
RT "PAX8 mutations associated with congenital hypothyroidism caused by
RT thyroid dysgenesis.";
RL Nat. Genet. 19:83-86(1998).
RN 161;
RP VARIANT ID TYR-57;
EX MEDLINE=11342381; PubMed 11232006;
PA Villain C., Pydlowski C., Dupre L., Heinrichs C., Abramowicz M.,
PA Malvaux P., Remenbeck B., Parpa J., Costagliola S., Vassart G.;
RT "Autosomal dominant transmission of congenital thyroid hypoplasia due
RT to loss-of-function mutation of PAX8.";
RL J. Clin. Endocrinol. Metab. 85:234-238(2001).
CC 1; FUNCTION: Transcription factor for the thyroid specific
CC expression of the genes exclusively expressed in the thyroid cell
CC type, maintaining the functional differentiation of such cells.
CC 1; SUBCELLULAR LOCATION: Nuclear.
CC 1; ALTERNATIVE PROCTS: 5 isoforms, 1-55, 2-55, 3-55, 4-55,
CC 4-Pax8 and 5-Pax8v; are produced by alternative splicing.
CC 1; TISSUE SPECIFICITY: EXPRESSED IN THE EXCRETORY SYSTEM, THYROID
CC GLAND AND WILMS TUMORS.
CC 1; DEVELOPMENTAL STAGE: IN DEVELOPING EXCRETORY SYSTEM, DURING
CC THYROID DIFFERENTIATION AND IN ADULT THYROID.
CC 1; DISEASE: Defects in PAX8 are a cause of thyroid dysgenesis (TD),
CC the most frequent cause of congenital hypothyroidism (CH). The
CC thyroid gland can be absent (athyreosis), ectopically located
CC (most often at the base of the tongue) and/or severely reduced in
CC size (hypoplasia).
CC 1; SIMILARITY: CONTAINS 1 PAIRED BOX DOMAIN.
CC
CC THIS SWISS PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.ebi.ac.uk/ebc/
CC or send an email to license@ebi.ac.uk).
CC
CC EMBL: X69699; -- NOT ANNOTATED_CDS.
CC EMBL: L19606; AAA03539.1; --
CC EMBL: S77904; AAB34216.1; --
CC EMBL: S77905; AAB34217.2; --
CC EMBL: S77406; AAB34218.2; --
CC FRL: A49126; A49126.
CC HSSP: P26367; 6PAX.
CC TRANSFAC: T01824; --
CC TRANSFAC: T01825; --
CC TRANSFAC: T01826; --
CC TRANSFAC: T01827; --
CC TRANSFAC: T02898; --
CC GeneW: HRPN-0622; PAX8.
CC MIM: 167415; --
CC MIM: 218700; --
CC INTERPRO: IPR001523; Paired_box.

```





FT TRANSMEM 6 26 POTENTIAL  
 FT DOMAIN 547 758 HISTIDINE KINASE  
 FT DOMAIN 9 421 ALA-RICH  
 FT MOD.RES 550 550 PHOSPHORYLATION (AUTO-)  
 FT CONFLICT 200 200 Q -> H (IN REF. 1)  
 FT CONFLICT 216 216 V -> E (IN REF. 1)  
 SQ SEQUENCE 769 AA: 82796 MW: 002B2428F18A57EF CRC64:  
  
 Query Match 4.58; Score 103; DB 1; Length 769;  
 Rest Local Similarity 22.98; Pred. No. 1.4;  
 Matches 91; Conservative 32; Mismatches 144; Indels 140; Gaps 19;  
  
 QY 46 KAIQL- MDVTPDAWPTWPKVIGELT ---PRGGELIAYLGHYWRQL----- 88  
 DB 345 KAALARYEDHQAHLKWLDPGRGLKVVRRPHPLGGMLLYSDITGRLRLKAQYNALI 404  
 QY 89 VAKRLIKKCGPQSGQVATADVDERTKRTGEAFAA-----GLAPDCA 131  
 DB 405 QVQQAIIILKL-----NDAVAVGS-DGRLRLINEAFETENNVTPHIAEAAGDFEGVVELCV 459  
 QY 132 LIVH-----TQADISSHDLF-----NPKIGVCULDNAVTD 164  
 DB 460 PRHLDSWRRLKGVADPTQMRAFTSGEVTSISRIVLQSRKLPDRAITLIAPADVD 519  
 QY 165 ALLERAGSIADPTGHYQTAPEFELFERNFQSNIAFKPEKQDESQSLTQALPSELKVSA 224  
 DB 520 T-----NDLQSAI-----ADRSAAIAFA-----ERLKR 542  
 QY 225 DVVSTIGAVS--LASMLTEIF-----LLQAQMPDPGWRITDSQWNTLIS-----IH 272  
 DB 543 DFP -GVSYELRTPLETTIGYSLELERAAGTISERGENHVAAVRAAAIQLARSTDDVLD 599  
 QY 273 NAQFELAGRTPEVAKSRAPHLLELILKIALDPPTPKQAVGVTLPTSVLFIACHDTNLANI 332  
 DB 600 MAQIDAGEMALETEDRYS--DLLNAQPRALKDAQLGGVTLAVE-----CEDVGLIR- 651  
 QY 333 GALELNNWILPGQPIN-----TPPGELVFERWRRLSD 365  
 DB 652 GDERKGAUTLDHLVHNAIRQITPGGRVTLSSARRALGE 688

Search completed: December 13, 2002, 15:54:04  
 Job time : 15 secs



Db 401 PAT STNSQWVSVLVEQTLQCMRKPRTSLNTPPEVRLTSLAAGVFNRAQCMCSLAGFTQ 420

QY 421 IVNEARIPACSLRSH 435

|||||

Db 421 IVNEARIPACSLRSH 435

RESULT 2

US-09-418-528-2

; Sequence 2, Application US/09-418528

; Patent No. 6183740

; GENERAL INFORMATION:

; APPLICANT: Kretz, Keith

; TITLE OF INVENTION: NOVEL PHYTASE

; FILE REFERENCE: 09010/029003

; CURRENT APPLICATION NUMBER: 09/09-019,598

; CURRENT FILING DATE: 1999-05-25

; EARLIER APPLICATION NUMBER: 09/291,931

; EARLIER FILING DATE: 1999-04-13

; EARLIER APPLICATION NUMBER: 08/910,798

; EARLIER FILING DATE: 1997-08-13

; EARLIER APPLICATION NUMBER: 09/259,214

; EARLIER FILING DATE: 1999-03-01

; NUMBER OF SEQ ID NOS: 4

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 2

; LENGTH: 440

; TYPE: PRT

; ORGANISM: Escherichia coli

US-09-418-528-2

Query Match 96.0%; Score 2188; Db 4; Length 440;

Best Local Similarity 97.7%; Prod. No. 1,70,222;

Matches 425; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1 MKAILIPFELSLIPITPSAFAUSEPELKLESVVIVSRBVRKATIKATQIMQIVADAMP 60

|||||

Db 1 MKAILIPFELSLIPITPSAFAUSEPELKLESVVIVSRBVRKATIKATQIMQIVADAMP 60

QY 61 TPVVKLSLTPRGELIAYLGHYWRQRLVADKLLKPGQPCQSQVAIADVDFRPKTGE 120

|||||

Db 61 TPVVKLSLTPRGELIAYLGHYWRQRLVADKLLKPGQPCQSQVAIADVDFRPKTGE 120

QY 121 AFAAGLAPDAATVHTQADTSSDDEPFIKRVGQGMANVTDAIIPFAGGSLAPFTCH 180

|||||

Db 121 AFAAGLAPDAATVHTQADTSSDDEPFIKRVGQGMANVTDAIIPFAGGSLAPFTCH 180

QY 181 YQTAFFRELERVLPQSNLTGKPKQDQSNLSLQALPSLSKVSAPVSLDIAVSLASMLT 240

|||||

Db 181 YQTAFFRELERVLPQSNLTGKPKQDQSNLSLQALPSLSKVSAPVSLDIAVSLASMLT 240

QY 241 EFLIQAGQMPDQWGRITDSHWNLTLSLHNAQFYLLQPTFIVARSRATPLDLIKTA 300

|||||

Db 241 EFLIQAGQMPDQWGRITDSHWNLTLSLHNAQFYLLQPTFIVARSRATPLDLIKTA 300

QY 401 LTPHPKQAGVTLPTSVLEIADHDTNLANIAGALELNWTLPGQDNTPPGGELVFERW 360

|||||

Db 401 LTPHPKQAGVTLPTSVLEIADHDTNLANIAGALELNWTLPGQDNTPPGGELVFERW 360

QY 461 RLRSNSQWLVSVLFTQTLQOMRKTPLSLNTPPGEVKLTLAGCPHNAQGMCSLAGFTQ 420

|||||

Db 461 RLRSNSQWLVSVLFTQTLQOMRKTPLSLNTPPGEVKLTLAGCPHNAQGMCSLAGFTQ 420

QY 421 IVNEARIPACSLRSH 435

|||||

Db 421 IVNEARIPACSLRSH 435

RESULT 3

US-09-291-941-2

; Sequence 2, Application US/09291931A

; Patent No. 6190897

; GENERAL INFORMATION:

; APPLICANT: Kretz, Keith

; TITLE OF INVENTION: NOVEL PHYTASE

; FILE REFERENCE: 09010/029003

; CURRENT APPLICATION NUMBER: US/09/291,931A

; CURRENT FILING DATE: 1999-04-13

; EARLIER APPLICATION NUMBER: 08/910,798

; EARLIER FILING DATE: 1997-08-13

; EARLIER APPLICATION NUMBER: 09/259,214

; EARLIER FILING DATE: 1999-03-01

; NUMBER OF SEQ ID NOS: 4

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 2

; LENGTH: 440

; TYPE: PRT

; ORGANISM: Escherichia coli

US-09-291-941-2

Query Match 96.0%; Score 2188; Db 4; Length 440;

Best Local Similarity 97.7%; Prod. No. 1,70,222;

Matches 425; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1 MKAILIPFELSLIPITPSAFAUSEPELKLESVVIVSRBVRKATIKATQIMQIVADAMP 60

|||||

Db 1 MKAILIPFELSLIPITPSAFAUSEPELKLESVVIVSRBVRKATIKATQIMQIVADAMP 60

QY 61 TPVVKLSLTPRGELIAYLGHYWRQRLVADKLLKPGQPCQSQVAIADVDFRPKTGE 120

|||||

Db 61 TPVVKLSLTPRGELIAYLGHYWRQRLVADKLLKPGQPCQSQVAIADVDFRPKTGE 120

QY 121 AFAAGLAPDAATVHTQADTSSDDEPFIKRVGQGMANVTDAIIPFAGGSLAPFTCH 180

|||||

Db 121 AFAAGLAPDAATVHTQADTSSDDEPFIKRVGQGMANVTDAIIPFAGGSLAPFTCH 180

QY 181 YQTAFFRELERVLPQSNLTGKPKQDQSNLSLQALPSLSKVSAPVSLDIAVSLASMLT 240

|||||

Db 181 YQTAFFRELERVLPQSNLTGKPKQDQSNLSLQALPSLSKVSAPVSLDIAVSLASMLT 240

QY 241 EFLIQAGQMPDQWGRITDSHWNLTLSLHNAQFYLLQPTFIVARSRATPLDLIKTA 300

|||||

Db 241 EFLIQAGQMPDQWGRITDSHWNLTLSLHNAQFYLLQPTFIVARSRATPLDLIKTA 300

QY 401 LTPHPKQAGVTLPTSVLEIADHDTNLANIAGALELNWTLPGQDNTPPGGELVFERW 360

|||||

Db 401 LTPHPKQAGVTLPTSVLEIADHDTNLANIAGALELNWTLPGQDNTPPGGELVFERW 360

QY 461 RLRSNSQWLVSVLFTQTLQOMRKTPLSLNTPPGEVKLTLAGCPHNAQGMCSLAGFTQ 420

|||||

Db 461 RLRSNSQWLVSVLFTQTLQOMRKTPLSLNTPPGEVKLTLAGCPHNAQGMCSLAGFTQ 420

QY 421 IVNEARIPACSLRSH 435

|||||

Db 421 IVNEARIPACSLRSH 435

RESULT 4

US-08-910-798-2

; Sequence 2, Application US/08910798

; Patent No. 5876997

; GENERAL INFORMATION:

; APPLICANT: KRETZ

; TITLE OF INVENTION: NOVEL PHYTASE

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fish & Richardson, P.C.

; STREET: 4225 Executive Square, Suite 1400

; CITY: La Jolla

; STATE: California

; COUNTRY: US

; ZIP: 92037

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

```

; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/910,799
; FILING DATE: August 13, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: HALL, PH.D., LISA A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/FEI NUMBER: 99016/029001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 423 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-910-798-2

```

```

Query Match          91.3%; Score 2080.5; DB 2; Length 423;
Best Local Similarity 91.8%; Pred. No. 3.5e-211;
Matches 408; Conservative 0; Mismatches 10; Indels 17; Gaps 1;

QY 1 MKALLPELSLLIPLTPQSAFAQSEPELKLESVVIVSRHGVRAPTKATQLMQDVTTPDAMP 60
DB 1 MKALLPELSLLIPLTPQSAFAQSEPELKLESVVIVSRHGVRAPTKATQLMQDVTTPDAMP 60

QY 61 TWPVKLGELTPRGSELIAYLGHYWPQRIVAIGSLIKKQSPQSQVALLADVDERTPKTGE 120
DB 61 TWPVKLGELTPRGSELIAYLGHYWPQRIVAIGSLIKKQSPQSQVALLADVDERTPKTGE 120

QY 121 AFAAGLAPDCAITVHTQADTSPPPLFNPLKTVGVGHQDNANVTDAILEPAGSIAFTFGH 180
DB 120 -----QADTSPPPLFNPLKTVGVGHQDNANVTDAILEPAGSIAFTFGH 180

QY 181 YOTAPRELRLVINEPQSNLCKRKQDESCSLTOALPSELKVSADCVSLTCAVSLASMLT 240
DB 164 KQTAPRELRLVINEPQSNLCKRKQDESCSLTOALPSELKVSADCVSLTCAVSLASMLT 223

QY 241 EFLQQAQGPFPQWGPRIETSHQNWILLSENAFELLLQKPEVAKSKATPFLULLIKTA 300
DB 224 EFLQQAQGPFPQWGPRIETSHQNWILLSENAFELLLQKPEVAKSKATPFLULLIKTA 283

QY 401 LTPHPQKQAYGVTLPTSVLFIAQHDNLNLAGALFNLTLPQGPUNTTPGGELVFERPW 360
DB 284 LTPHPQKQAYGVTLPTSVLFIAQHDNLNLAGALFNLTLPQGPUNTTPGGELVFERPW 343

QY 461 RLINSQWIGVSVLVTQLOQMPKPTPLSTNTPGSPVKLALAGPFPKNAQGMCSLAFITQ 420
DB 344 RLINSQWIGVSVLVTQLOQMPKPTPLSTNTPGSPVKLALAGPFPKNAQGMCSLAFITQ 403

QY 421 IVNEAPTPACSLPSH 435
DB 404 IVNEAPTPACSLPSH 418

```

## RESULT 5

```

US-09-044-718-3
; Sequence 3, Application US/09044718
; Patent No. 6391605
; GENERAL INFORMATION:
; APPLICANT: KOSTREWA, Dirk
; APPLICANT: PASAMONTES, Luis
; APPLICANT: TOMSCHY, Andrea
; APPLICANT: van LOON, Adolphus
; APPLICANT: VOGEL, Kurt
; APPLICANT: WYSS, Markus
; TITLE OF INVENTION: MODIFIED PHYTASES
; FILE REFERENCE: Modified Phytases
; CURRENT APPLICATION NUMBER: US/09/044,718
; CURRENT FILING DATE: 1998-03-19
; PRIOR APPLICATION NUMBER: EP 97810175.6

```

```

; PRIOR FILING DATE: 1997-03-25
; NUMBER OF SEQ ID NOS: 82
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 439
; TYPE: PRT
; ORGANISM: Aspergillus fumigatus
US-09-044-718-3

```

```

Query Match          5.5%; Score 126; DB 4; Length 449;
Best Local Similarity 20.9%; Pred. No. 0.00012;
Matches 102; Conservative 69; Mismatches 169; Indels 199; Gaps 26;

QY 7 PELSLLIPLTPQSAFAQSEPELKLESVVIVSRHGVRAPTKATQLMQDVTTPDAMP 58
DB 26 PELSLEDELSVSSKLPK---DGRITLVQVLSRHGARYETTSKSKYKKLVIAIQANATDF 82

QY 59 -----WPTWPKLG--ELTPRGSELIAYLGHYWPQRIVAIGSLIKKQSPQSQVALL 108
DB 83 KQKFAFLKTYNYTLGADDLTPFGROLYNSGIKIFYORYKALARSVVP-----FIR 132

QY 109 ADVDERTPKTGEAFAAGL-----APDCAITVHIALISSDPLFNPLKTV 154
DB 133 ASSGDFVIASGEKFTGEGQAAKLADPGATNKAAPALSVLI--PESEIFN-----NILDHGV 186

QY 155 C-OLDNANVTDAILEPAGSIAFTFGHYGTAPFTEFEVNFQGSNLCIKPKQDESCSLT 214
DB 187 CTKFAASQLGDEV-----AANFTALFADIR-----AKAEKII----- 218

QY 214 QALPSELKVSADCVSLTCAVSLASMLTETETLQQAQGPFPQWGRITNSHOWNLILSL-- 271
DB 219 --LPGVILIDEDVSLMDMCSFDTV-----ARTSDASQLSPFQQLPT 258

QY 272 HN--AQFDLLQR-----TPE-VARSATPELDDLIKT--ALTPHP 305
DB 259 HNEWKYNYLQSLGKYRYGACGNLPQAGTGFTNFIAPITRSPVQDHTSTNLSVSNP 318

QY 306 PQKQAYGVTLPTSVLFTIA--GHDTNLANIAGALFNLTLPQGPUNTTPGGELVFERWRKL 363
DB 319 -----ATPPLNATMYVDVSHDMSVSIFFAL-----GIYNTEPLSPISVESAKEL 364

QY 364 SD-NSQW1---QVSLVFTQLOQMPKPTPLSTNTPGSPVKLTIAGPFPKNAQGMCSLAFIT 419
DB 365 DGYSSAVVPPGARAYFETMQCKSEKPL-VRALINDKVVPLRGNDV-DKLGRCCKINDV 422

QY 420 QIVNEAR 426
DB 423 KQLSWAR 429

RESULT 6
US-09-044-718-12
; Sequence 12, Application US/09044718
; Patent No. 6391605
; GENERAL INFORMATION:
; APPLICANT: KOSTREWA, Dirk
; APPLICANT: PASAMONTES, Luis
; APPLICANT: TOMSCHY, Andrea
; APPLICANT: van LOON, Adolphus
; APPLICANT: VOGEL, Kurt
; APPLICANT: WYSS, Markus
; TITLE OF INVENTION: MODIFIED PHYTASES
; FILE REFERENCE: Modified Phytases
; CURRENT APPLICATION NUMBER: US/09/044,718
; CURRENT FILING DATE: 1998-03-19
; PRIOR APPLICATION NUMBER: EP 97810175.6
; PRIOR FILING DATE: 1997-03-25
; NUMBER OF SEQ ID NOS: 82
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 12
; LENGTH: 449
; TYPE: PRT
; ORGANISM: Aspergillus fumigatus

```

US 09-044-718 12

Query Match 5.5%; Score 126; ID 4; Length 449;

Best local Similarity 20.9%; Pred. No. 0.00013;

Matches 102; Conservative 69; Mismatches 166; Indels 150; Gaps 26;

QY 7 PFLSLPLTPQSAFAQSEPELKLESVIVSRHCVKAPT-----KATQMQADVTIDVA--- 58  
 DB 46 PFLSLPLTPQSAFAQSEPELKLESVIVSRHCVKAPT-----KATQMQADVTIDVA--- 58  
 QY 59 WPTWPKVKG--ELTPRGDELLAYLG--HYWQRLVAQGLLTKCCGQSGOVAVI 108  
 DB 93 KGEAFIKTYNYTLGALDLPFGEGQVLVNSGIRFYQYKALASVVP-----FIR 142  
 QY 109 AWDETRTKGEAFAGL-----APKCATIVHTQADTSSEPELNPDKGV 174  
 DB 143 ASGSDVIVASGKRFIEGPGQAKLADPGATNKAAPAVSVII PESETFN NILLHGV 212  
 QY 155 C-OLDNANVTDAILERAGGSTAIDTGHYQTAFLERELVNFQSNLCLKPKQJESCSLT 214  
 DB 197 CTKEASQGLDEV-----AANFALFADIR-----ARAERH----- 228  
 QY 214 QALPSELKVSADGVSLTGAVSLASMLTETFLQQAQGMPEFWGRTIUSHQWNLISL 271  
 DB 229 --LPGVLTIDEDVVSLSMOMSEFIV-----TPF VAKSRATPFLDLKLT--ALPHP 305  
 QY 272 HN--AOFLLQR-----TPF VAKSRATPFLDLKLT--ALPHP 305  
 DB 269 HNWKKRYNYLQSGIKYGYGAGNPLGAGTGFTNLLIAGLQSPVLDHISTINSTVSNP 328  
 QY 406 PQQARGVTLTPTSVLEFIA--GHDTNLANLGGALELNKTLPGQDNPFGGELVFERWRL 363  
 DB 329 --ATEPLNATMYVDESHUNSVSIFPAL-----GIYNGTEPLSPTSVESAKEL 374  
 QY 464 SD--NSQWT--QVSLVQTLQAMRKDTLSINIPGQKVLITAGCTERNAGQMGSLAGFT 419  
 DB 475 DYSASVWVPEGAPAYETMQLKSKPEPL--VFALINRVVPLRGCEV--DKLGCKLNDEV 442  
 QY 420 QIVNEAR 426  
 DB 433 KGLSWAR 439

RESULT 7

US 08-868-435 34

Sequence 34, Application US/08068435

Patent No. 6291221

GENERAL INFORMATION:

APPLICANT: Van Loon, Adolphus

TITLE OF INVENTION: POLYPEPTIDES WITH PHYTASE ACTIVITY

NUMBER OF SEQUENCES: 35

CORRESPONDENCE ADDRESS:

ADDRESSEE: Hoffmann-La Roche Inc.

STREET: 400 Kingsland Street

CITY: Nutley

STATE: New Jersey

COUNTRY: United States of America

ZIP: 07110

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APP. FILING NUMBER: 08/090,435

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/744,241

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Kass, Alan P

REGISTRATION NUMBER: 42142

REFERENCE/BOOK NUMBER: Case booked 9409

TELECOMMUNICATION INFORMATION:

TELEPHONE: (201) 245-4205

TELEFAX: (201) 245-2464

INFORMATION FOR SEQ ID NO: 34:

SEQUENCE CHARACTERISTICS:

LENGTH: 465 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

FEATURES:

NAME/KEY: misc\_feature

LOCATION: 104

OTHER INFORMATION: /note "potential N glycosylation site"

FEATURES:

NAME/KEY: misc\_feature

LOCATION: 119

OTHER INFORMATION: /note "potential N glycosylation site"

FEATURES:

NAME/KEY: misc\_feature

LOCATION: 205

OTHER INFORMATION: /note "potential N glycosylation site"

FEATURES:

NAME/KEY: misc\_feature

LOCATION: 228

OTHER INFORMATION: /note "potential N glycosylation site"

FEATURES:

NAME/KEY: misc\_feature

LOCATION: 347

OTHER INFORMATION: /note "potential N glycosylation site"

FEATURES:

NAME/KEY: misc\_feature

LOCATION: 374

OTHER INFORMATION: /note "potential N glycosylation site"

US-08-868-435-34

Query Match

Best local Similarity 20.9%; Score 126; ID 4; Length 465;

Matches 102; Conservative 69; Mismatches 166; Indels 150; Gaps 26;

QY 7 PFLSLPLTPQSAFAQSEPELKLESVIVSRHCVKAPT-----KATQMQADVTIDVA--- 58  
 DB 52 PFLSLPLTPQSAFAQSEPELKLESVIVSRHCVKAPT-----KATQMQADVTIDVA--- 58  
 QY 59 WPTWPKVKG--ELTPRGDELLAYLG--HYWQRLVAQGLLTKCCGQSGOVAVI 108  
 DB 109 KGEAFIKTYNYTLGALDLPFGEGQVLVNSGIRFYQYKALASVVP-----FIR 142  
 QY 109 AWDETRTKGEAFAGL-----APKCATIVHTQADTSSEPELNPDKGV 174  
 DB 159 ASGSDVIVASGKRFIEGPGQAKLADPGATNKAAPAVSVII PESETFN NILLHGV 212  
 QY 155 C-OLDNANVTDAILERAGGSTAIDTGHYQTAFLERELVNFQSNLCLKPKQJESCSLT 214  
 DB 213 CTKEASQGLDEV-----AANFALFADIR-----ARAERH----- 244  
 QY 214 QALPSELKVSADGVSLTGAVSLASMLTETFLQQAQGMPEFWGRTIUSHQWNLISL 271  
 DB 245 --LPGVLTIDEDVVSLSMOMSEFIV-----TPF VAKSRATPFLDLKLT--ALPHP 305  
 QY 272 HN--AOFLLQR-----TPF VAKSRATPFLDLKLT--ALPHP 305  
 DB 285 HNWKKRYNYLQSGIKYGYGAGNPLGAGTGFTNLLIAGLQSPVLDHISTINSTVSNP 328  
 QY 406 PQQARGVTLTPTSVLEFIA--GHDTNLANLGGALELNKTLPGQDNPFGGELVFERWRL 363  
 DB 445 --ATEPLNATMYVDESHUNSVSIFPAL-----GIYNGTEPLSPTSVESAKEL 374  
 QY 464 SD--NSQWT--QVSLVQTLQAMRKDTLSINIPGQKVLITAGCTERNAGQMGSLAGFT 419  
 DB 475 DYSASVWVPEGAPAYETMQLKSKPEPL--VFALINRVVPLRGCEV--DKLGCKLNDEV 442







```
Db 245 --LPRVILHEDVVSIMDMCSFDIV-----ARTSDASOLSFCOLFT 284
QY 272 HN--ACPNLLR-----TPF-VAPSPATPLDLIKT--ALRHP 305
Db 285 HNEWKYNYLSLCKYGYGAGNPIGPAQIGFTNFIARITSPQDHSTNSTLSNP 344
QY 306 PQKQAYCVLTPTSVLFA--GHVNLNLANLGGAI FNNWTLPCQDNTDGGPHVPEPWPPI 363
Db 345 -----ATEPLNATMYVDFSHDSNMSVSIFFAL-----GLYNGTELSPTSIVESAKEL 390
QY 364 SD--NSQWI---QVSLVFATLQMPDKTPLSLNTPPEVFKITIAACEFPNAGQMSLACFT 419
Db 391 DQYSASWVWVPGARAYFETMCKSEKKEPL-VKALLINRVVPLRGCTV-DKLRGKLNDFV 448
QY 420 QVNEAP 426
Db 449 KGLSWAR 455

RESULT 12
US-09-044-718-81
: Sequence 81, Application US/09044718
: Patent No. 6391605
: GENERAL INFORMATION:
: APPLICANT: KOSTERWA, Dirk
: APPLICANT: PASAMONTES, Luis
: APPLICANT: TOMSCHY, Andrea
: APPLICANT: van LOON, Adolphus
: APPLICANT: VOGEL, Kurt
: APPLICANT: WYSS, Markus
: TITLE OF INVENTION: MODIFIED PHYTASES
: FILE REFERENCE: Modified phytases
: CURRENT APPLICATION NUMBER: US09/044,718
: CURRENT FILING DATE: 1998-03-19
: PRIOR APPLICATION NUMBER: FF 97810175 6
: PRIOR FILING DATE: 1997-03-25
: NUMBER OF SEQ ID NOS: 82
: SOFTWARE: Patent In Ver. 2.1
: SEQ ID NO 81
: LENGTH: 465
: TYPE: PRT
: ORGANISM: Aspergillus fumigatus
US-09-044-718-81

Query Match 5.18; Score 116; DB 4; Length 465;
Best Local Similarity 29.7%; Pred. No. 0.0015;
Matches 101; Conservative 50; Mismatches 167; Indels 150; Gaps 26;

QY 7 PPLSLILPLTPQSAFASDEPEKLESVVIVSHVHPAPT-----KATQIMQDVTPDA--- 58
Db 52 PHSLEGEELSVSKLFE---DPRITLVVLSFHFAPYFTSSKSKYKLVTAIGANATEF 108
QY 59 -----WPTWPKLG--SLTPGSGELIAYIG--HYWPGPIVADQILPKCGTPQSQVAIL 108
Db 109 KRLALALDYD LGAAGALAL ELQAQVHSEITFFVGYKALAKRSVVP--- FIR 159
QY 159 ADVPEKTRKTKTGTAAGL ..... AFDCATTVHTQADTSPPDLPLNLTGV 154
Db 159 ASGSDRVITASEKFTIEGQQAKLADPGATNRAAPISVLI-PSETFN-----NTLDHGV 212
QY 175 Q QLPANVWDALEPAAGS-ADTCHYLAFFLEPVI NPDQSNLTJRKPKQRESQSLT 213
Db 213 CTKFASQIGDEV-----AANATALEAPDIR-----ARAKKH----- 244
QY 214 QALPSELKVSADCVSLGTAGSLASMLTEFLIQAGSGMPFGWGRITDSHOWNTLLSL- 271
Db 245 LFCVTLTDHVDVSLDMCSFDIV-----APTSASQSLSPFQULPT 284
QY 272 HN--ACPNLLR-----TPF-VAPSPATPLDLIKT--ALRHP 305
Db 285 HNEWKYNYLSLCKYGYGAGNPLGPAJGIGFTNFIARITSPQDHSTNSTLSNP 344
QY 306 PQKQAYCVLTPTSVLFA--GHVNLNLANLGGAI FNNWTLPCQDNTDGGPHVPEPWPPI 363
```

```
Db 345 -----ATEPLNATMYVDFSHDSNMSVSIFFAL-----GLYNGTELSPTSIVESAKEL 390
QY 364 SD--NSQWI---QVSLVFATLQMPDKTPLSLNTPPEVFKITIAACEFPNAGQMSLACFT 419
Db 391 DQYSASWVWVPGARAYFETMCKSEKKEPL-VKALLINRVVPLRGCTV-DKLRGKLNDFV 448
QY 420 QVNEAP 426
Db 449 KGLSWAR 455

RESULT 13
US-07-923-724-2
: Sequence 2, Application US/07923724
: Patent No. 5780292
: GENERAL INFORMATION:
: APPLICANT: Nevalainen, Helena K.M.
: APPLICANT: Paloheimo, Marja T.
: APPLICANT: Miettinen-Oinonen, Arja S.K.
: APPLICANT: Torkkeli, Tuula K.
: APPLICANT: Cantrell, Michael
: APPLICANT: Piddington, Christopher S.
: APPLICANT: Rambesek, Joha A.
: APPLICANT: Turunen, Marja K.
: APPLICANT: Fagerstr m, Richard H.
: TITLE OF INVENTION: Production of Phytase Degrading Enzymes
: NUMBER OF SEQUENCES: 66
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Stetner, Kessler, Goldstein & Fox
: STREET: 1100 New York Avenue, Suite 600
: CITY: Washington
: STATE: D.C.
: COUNTRY: U.S.A
: ZIP: 20005
: COMPUTER PROGRAM: FORM
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/923,724
: FILING DATE: 31-JUL-1992
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 97/496,155
: FILING DATE: 19-MAR-1990
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/044,077
: FILING DATE: 29-APR-1987
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: UK 8610600
: FILING DATE: 30-APR-1985
: ATTORNEY/AGENT INFORMATION:
: NAME: Cimbala, Michele A.
: PRACTICE NUMBER: 33,851
: REFERENCE/DOCKET NUMBER: 1050.0240004
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202) 371-2600
: TELEFAX: (202) 371-2540
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 479 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-07-923-724-2

Query Match 4.88; Score 109.5; DB 1; Length 479;
Best Local Similarity 19.2%; Pred. No. 0.0078;
Matches 91; Conservative 48; Mismatches 145; Indels 189; Gaps 23;
```

```

QY 29 KLESVIVSRHGVAPTKATGLMDGVTPDAMPWPVKL-----GFLT----- 70
DB 72 EVQVIMVKHGRKRYSPSAG--KIDGFAIAKYSINLFFYKGLADLAINWTVVYPNFCY 129
QY 71 -----PRGGELIAY-----LGHYWRQRLVAIKLIPKGGCHQSQVAILADWDE 114
DB 130 YNAETTSQVAGLLDAYNHCNKKARYGHLWNGETVVEF-----SSGYGRVI-----E 178
QY 114 KTKTCTCA--TAAGLAPICATVTHQATSTSPDIPRIKTCVQGLIHANVTDAILLPAGG 172
DB 179 TARKGGEGFYNNSTNAALNITSESEVMGADSL-----TPTCDTNDQITCDNL----- 229
QY 173 STADFTGHVCTAFRELERLNEFNSICLKE--EKQESLSITQALFSELKVSACVSLIG 231
DB 230 -----YQ-----LPGKVAAPAINSONPMNLT-----ASIVYNL----- 259
QY 232 AVSLASMLTEFLLOAQGMPFGWGRITDSHOWNTL-----LSLH----- 272
DB 260 MYWASF-----ELNAPFESNWINAFTQDEWVSFGYVEDLNYYCAGPGUKNMAAVG 310
QY 273 -----NAOFDILQRTPEVARSKATFLDLIKTALTPHPQKAYGVTLPTISVLFIAGHDIN 328
DB 311 AVYANASLTLNQPKA-----GSLFFNFADTN 340
QY 329 LANIGGALEL---NWTLF-----GQP-----DNTPPGGELVFERW-----RLSDNSQWIQ 371
DB 341 TITPILAAAGVLIPNEDLPDRVAFGNPYSIGNIVPMGGHLLTIERLSQATALSDEGTVVR 400
QY 372 VSLVEQTLDEMBOKTPLSLNTPPEVWLTLAGYFERNAGCMCSLACHTQIVNE 424
DB 401 LVL-----NEAVLPNDTISGPNYS--CPLANYSTLNLK 432

```

## RESULT 14

```

US-08-609-426A-2
Sequence 4: Application US/08/426A-2
Patent No. 5840733
GENERAL INFORMATION:
APPLICANT: NEVALAINEN, HELENA K.M.
APPLICANT: PALOHEIMO, MARJA T.
APPLICANT: MIETTINEN-OININEN, ARJA S.K.
APPLICANT: TORKKELI, TOULU K.
APPLICANT: CANTRELL, MICHAEL
APPLICANT: PIDDINGTON, CHRISTOPHER S.
APPLICANT: RAMHOSSEK, JOHN A.
APPLICANT: FAGERSTROM, RICHARD B.
APPLICANT: TURONEN, MARJA K.
TITLE OF INVENTION: Production of Phytase Degrading Enzymes
TITLE OF INVENTION: In Trichoderma
NUMBER OF SEQUENCES: 69
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox
STREET: 1100 New York Avenue, Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: MS-DOS/MS-WIN
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/609,426A
FILING DATE: 01-MAR-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/923,724
FILING DATE: 31-JUL-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/496,155
FILING DATE: 19-MAR-1990

```

```

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/044,077
FILING DATE: 29-APR-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: UK 8610600
FILING DATE: 30-APR-1986
ATTORNEY/AGENT INFORMATION:
NAME: Reed, Grant E.
REGISTRATION NUMBER: P-41,264
REFERENCE/DOCKET NUMBER: 1050,000001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 471-2600
TELEFAX: (202) 471-2540
INFORMATION FOR SEQ. ID NO. 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 479 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-609-426A-2
Query Match 4.88; Score 109.5; DB 2: Locals, 479;
Best Local Similarity 19.28; Prod. No. 010076;
Matches 91; Conservative 48; Mismatches 145; Indels 189; Gaps 24;
QY 29 KLESVIVSRHGVAPTKATGLMDGVTPDAMPWPVKL-----GFLT----- 70
DB 72 EVQVIMVKHGRKRYSPSAG--KIDGFAIAKYSINLFFYKGLADLAINWTVVYPNFCY 129
QY 71 -----PRGGELIAY-----LGHYWRQRLVAIKLIPKGGCHQSQVAILADWDE 114
DB 130 YNAETTSQVAGLLDAYNHCNKKARYGHLWNGETVVEF-----SSGYGRVI-----E 178
QY 114 KTKTCTCA--TAAGLAPICATVTHQATSTSPDIPRIKTCVQGLIHANVTDAILLPAGG 172
DB 179 TARKGGEGFYNNSTNAALNITSESEVMGADSL-----TPTCDTNDQITCDNL----- 229
QY 173 STADFTGHVCTAFRELERLNEFNSICLKE--EKQESLSITQALFSELKVSACVSLIG 231
DB 230 -----YQ-----LPGKVAAPAINSONPMNLT-----ASIVYNL----- 259
QY 232 AVSLASMLTEFLLOAQGMPFGWGRITDSHOWNTL-----LSLH----- 272
DB 260 MYWASF-----ELNAPFESNWINAFTQDEWVSFGYVEDLNYYCAGPGUKNMAAVG 310
QY 273 -----NAOFDILQRTPEVARSKATFLDLIKTALTPHPQKAYGVTLPTISVLFIAGHDIN 328
DB 311 AVYANASLTLNQPKA-----GSLFFNFADTN 340
QY 329 LANIGGALEL---NWTLF-----GQP-----DNTPPGGELVFERW-----RLSDNSQWIQ 371
DB 341 TITPILAAAGVLIPNEDLPDRVAFGNPYSIGNIVPMGGHLLTIERLSQATALSDEGTVVR 400
QY 372 VSLVEQTLDEMBOKTPLSLNTPPEVWLTLAGYFERNAGCMCSLACHTQIVNE 424
DB 401 LVL-----NEAVLPNDTISGPNYS--CPLANYSTLNLK 432

```

## RESULT 15

```

US-08-374-652C-4
Sequence 4: Application US/08/474652C
Patent No. 584286
GENERAL INFORMATION:
APPLICANT: NEVALAINEN, HELENA K.M.
APPLICANT: PALOHEIMO, MARJA T.
APPLICANT: FAGERSTROM, RICHARD B.
APPLICANT: MIETTINEN-OININEN, ARJA S.
APPLICANT: TURONEN, MARJA K.
APPLICANT: RAMHOSSEK, JOHN A.
APPLICANT: PIDDINGTON, CHRISTOPHER S.
APPLICANT: HOUSTON, CHRISTINE S.
APPLICANT: CANTRELL, MICHAEL A.
TITLE OF INVENTION: RECOMBINANT CELLS, DNA CONSTRUCTS,

```



100

101

102

103

104

105

106

107

108

109

110

111

112

113

114

115

116

117

118

119

120

121

122

123

124

125

126

127

128

129

130

131

132

133

134

135

136

137

138

139

140

141

142

143

144

145

146

147

148

149

150

151

152

153

154

155

156

157

158

159

160

161

162

163

164

165

166

167

168

169

170

171

172

173

174

175

176

177

178

179

180

181

182

183

184

185

186

187

188

189

190

191

192

193

194

195

196

197

198

199

200

201

202

203

204

205

206

207

208

209

210

211

212

213

214

215

216

217

218

219

220

221

222

223

224

225

226

227

228

229

230

231

232

233

234

235

236

237

238

239

240

241

242

243

244

245

246

247

248

249

250

251

252

253

254

255

256

257

258

259

260

261

262

263

264

265

266

267

268

269

270

271

272

273

274

275

276

277

278

279

280

281

282

283

284

285

286

287

288

289

290

291

292

293

294

295

296

297

298

299

300

301

302

303

304

305

306

307

308

309

310

311

312

313

314

315

316

317

318

319

320

321

322

323

324

325

326

327

328

329

330

331

332

333

334

335

336

337

338

339

340

341

342

343

344

345

346

347

348

349

350

351

352

353

354

355

356

357

358

359

360

361

362

363

364

365

366

367

368

369

370

371

372

373

374

375

376

377

378

379

380

381

382

383

384

385

386

387

388

389

390

391

392

393

394

395

396

397

398

399

400

401

402

403

404

405

406

407

408

409

410

411

412

413

414

415

416

417

418

419

420

421

422

423

424

425

426

427

428

429

430

431

432

433

434

435

436

437

438

439

440

441

442

443

444

445

446

447

448

449

450

451

452

453

454

455

456

457

458

459

460

461

462

463

464

465

466

467

468

469

470

471

472

473

474

475

476

477

478

479

480

481

482

483

484

485

486

487

488

489

490

491

492

493

494

495

496

497

498

499

500

501

502

503

504

505

506

507

508

509

510

511

512

513

514

515

516

517

518

519

520

521

522

523

524

525

526

527

528

529

530

531

532

533

534

535

536

537

538

539

540

541

542

543

544

545

546

547

548

549

550

551

552

553

554

555

556

557

558

559

560

561

562

563

564

565

566

567

568

569

570

571

572

573

574

575

576

577

578

579

580

581

582

583

584

585

586

587

588

589

590

591

592

593

594

595

596

597

598

599

600

601

602

603

604

605

606

607

608

609

610

611

612

613

614

615

616

617

618

619

620

621

622

623

624

625

626

627

628

629

630

631

632

633

634

635

636

637

638

639

640

641

642

643

644

645

646

647

648

649

650

651

652

653

654

655

656

657

658

659

660

661

662

663

664

665

666

667

668

669

670

671

672

673

674

675

676

677

678

679

680

681

682

683

684

685

686

687

688

689

690

691

692

693

694

695

696

697

698

699

700

701

702

703

704

705

706

707

708

709

710

711

712

713

714

715

716

717

718

719

720

721

722

723

724

725

726

727

728

729

730

731

732

733

734

735

736

737

738

739

740

741

742

743

744

745

746

747

748

749

750

751

752

753

754

755

756

757

758

759

760

761

762

763

764

765

766

767

768

769

770

771

772

773

774

775

776

777

778

779

780

781

782

783

784

785

786

787

788

789

790

791

792

793

794

795

796

797

798

799

800

801

802

803

804

805

806

807

808

809

810

811

812

813

814

815

816

817

818

819

820

821

822

823

824

825

826

827

828

829

830

831

832

833

834

835

836

837

838

839

840

841

842

843

844

845

846

847

848

849

850

851

852

853

854

855

856

857

858

859

860

861

862

863

864

865

866

867

868

869

870

871

872

873

874

875

876

877

878

879

880

881

882

883

884

885

886

887

888

889

890

891

892

893

894

895

896

897

898

899

900

901

902

903

904

905

906

907

908

909

910

911

912

913

914

915

916

917

918

919

920

921

922

923

924

925

926

927

928

929

930

931

932

933

934

935

936

937

938

939

940

941

942

943

944

945

946

947

948

949

950

951

952

953

954

955

956

957

958

959

960

961

962

963

964

965

966

967

968

969

970

971

972

973

974

975

976

977

978

979

980

981

982

983

984

985

986

987

988

989

990

991

992

993

994

995

996

997

998

999

1000







QY 301 CACTCTGGTCAGTCGCGATATATCTCTATCTGGACGAGCGTAACCGGIAAAACAGCGAA 360  
 DB 101 GlnSerGlyGluValAlaIleIleAlaAspValAspGluArgThrArgGlyThrGlyGln 120  
 QY 361 GCTTGGCGGGGGCTCTGCACTGACTGACTGCTCAATAAAGGTATACAGTCAAGTATAG 420  
 DB 121 AlaPheValAspLysLeuAlaProAspCysAlaIleThrValIleThrGlnAlaAspThr 140  
 QY 421 TCCACATCCGATCCGCTTATTAATGCTCTAATAAACTGGCGCTTGGCAATTCGATACGCG 480  
 DB 141 SerGlnPheAspThrLeuPheAspProLeuLysThrGlyValGlySerLysGluAsnAla 160  
 QY 481 AAGCTGAGTGAAGCATGCTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540  
 DB 161 AsnValThrAspAlaIleGlnSerArgAlaValLysSerLysAlaAspPheThrGlyLys 180  
 QY 541 TATCAAGCGCGCTTTGCTGCACTGCAATGAGGTGCTTAATTTCCGCAATCAACATGTGAC 600  
 DB 181 ArgLysThrAlaPheArgGlnLeuGlnLysArgValLeuAsnPheProGlnSerAsnLeuGly 200  
 QY 601 GTTAACCTGCAAAACAGACAGCAAAAGCTTTCATTAAAGCAGAGCGATTATGATGCAACT 660  
 DB 201 LeuLysArgGlnLysGlnAspLysSerGlySerLeuThrGlnAlaLeuProGlnSerGlnLeu 220  
 QY 661 AAGCTGAGCGCGCTTCTGCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 720  
 DB 221 LysValSerAlaAspAsnValSerLeuThrGlyAlaValSerLeuAlaSerMetLeuThr 240  
 QY 721 CAGATATTCT 780  
 DB 241 GluThrPheLeuLeuGlnGlnAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 260  
 QY 781 CATTTCAACCTGCAAAACAGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 840  
 DB 261 AspSerHisGlnLysPheAsnThrLeuLeuSerLeuLysAsnAlaGlnPheThrLeuLeuGln 280  
 QY 841 CTAATCTGACAGATCT 900  
 DB 281 ArgThrProGlnValAlaArgSerArgAlaThrProLeuLeuAspLeuLysThrAla 300  
 QY 901 TTGACATCT 960  
 DB 301 LeuThrProHisProGlnLysGlnAlaLysValThrLeuProThrSerValLeu 320  
 QY 961 TTATCCGCGCGACAGCATATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1020  
 DB 321 PheThrAlaGlyHisAspThrAspLeuAlaAspLeuGlyGlyAlaLeuGlnLysAsnThr 340  
 QY 1021 AAGCTGCGCGCTTCT 1080  
 DB 341 ThrLeuProGlyGlnProAspAsnThrProProGlyGlyGlnValPheGlnArgThr 360  
 QY 1081 GTCGCTTAAGTATTAATAGTATGATGATGATGATGATGATGATGATGATGATGATGATG 1140  
 DB 361 ArgArgLeuSerAspAsnSerGlnTrpIleGlnValSerLeuValPheGlnLeuGln 380  
 QY 1141 CAGATGCTGATGATAAGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200  
 DB 381 GlnMetArgAspLysThrProLeuSerLeuAsnThrProProGlyGlyValLysLeuThr 400  
 QY 1201 CTGCGTAAAGTATTAAGTATGATGATGATGATGATGATGATGATGATGATGATGATG 1260  
 DB 401 LeuAlaGlyCysGlnGlnArgAspAlaGlnGlyMetCysSerLeuAlaGlyPheThrGln 420  
 QY 1261 AAGCTCAATGAAGCACTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1320  
 DB 421 IleValAsnGluAlaArgIleProAlaCysSerLeu 432  
 RESULT 2  
 ID AGP\_EC01 STAN1AR0 PK1 413 AA.  
 AC P19926;

DT 01-FEB-1991 (Rel. 17, Created)  
 E1 01-FEB-1991 (Rel. 17, Last sequence update)  
 DE 16-OCT-2001 (Rel. 40, Last annotation update)  
 EE Glucose 1-phosphatase precursor (EC 3.1.3.18) (G1Pase).  
 GN AGP OR B1002.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 CC Escherichia.  
 OX NCHI\_Taxid:562;  
 PN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 21-31.  
 FC STRAIN K12;  
 RX MEDLINE=90130418; PubMed=2153660;  
 RA Pradel E., Marek C., Boquet P.L.;  
 RT "Nucleotide sequence and transcriptional analysis of the Escherichia  
 coli *agp* gene encoding periplasmic acid glucose 1-phosphatase";  
 RL J. Bacteriol. 172:802-807(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 FC STRAIN K12 / MG1655.  
 RX MEDLINE=97426617; PubMed=9278503;  
 RA Pfaffner K.P., Plunkett G. III, Bloch C.A., Perna N.L., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.W.,  
 RA Gwien Z., Davis N.W., Kittling M.A., Glick M.A., Plesch D.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12";  
 RL Science 277:1453-1474(1997).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 FC STRAIN K12;  
 RX MEDLINE=97061192; PubMed=8405232;  
 RA Oshida T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,  
 RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kusai K., Kashimoto K.,  
 RA Kimura S., Kitagawa M., Makioka K., Mitsuoka S., Miki T., Mizutani K.,  
 RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,  
 RA Sasaki G., Seki Y., Tagami H., Takamizawa K., Waki T., Yamamoto Y.,  
 RA Yanai M., Horiuchi T.;  
 RT "A 716 kb DNA sequence of the Escherichia coli K-12 genome  
 corresponding to the 12.7-28.0 min region on the linkage map";  
 RL DNA Res. 3:137-155(1996).  
 RN [4]  
 RP SEQUENCE OF 23-34.  
 FC STRAIN K12 / EMG2;  
 RX MEDLINE 97443975; PubMed 9298646;  
 RA Link A.J., Robinson K., Church G.M.;  
 RT "Computing the predicted and observed properties of proteins encoded  
 in the genome of Escherichia coli K-12";  
 FL Electrophoresis 19:1259-1312(1997).  
 CC -1- FUNCTION: ABSOLUTELY REQUIRED FOR THE GROWTH OF E. COLI IN A HIGH-  
 PHOSPHATE MEDIUM CONTAINING IS-1-P AS THE SOLE CARBON SOURCE.  
 CC -1- CATALYTIC ACTIVITY: D-glucose 1-phosphate + H(2)O -> D-glucose +  
 phosphate.  
 CC -1- ENZYME REGULATION: INDEPENDENT FROM INORGANIC PHOSPHATE  
 AVAILABILITY, AND APPARENTLY SUBMITTED TO CAIACELLITE REPRESSION,  
 IT IS POSITIVELY CONTROLLED BY CAMP AND THE CAMP RECEPTOR PROTEIN.  
 CC -1- SUBUNIT: HOMODIMER.  
 CC -1- SUBCELLULAR LOCATION: Periplasmic.  
 CC -1- MISCELLANEOUS: OPTIMALLY ACTIVE AROUND PH 4.  
 CC -1- SIMILARITY: BELONGS TO THE HISTIDINE ACID PHOSPHATASE FAMILY.  
 CC  
 CC This SWISS-Prot entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is not used for  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see http://www.ssi-b.ch/ or send an email to license@ssi-b.ch).  
 CC  
 DB EMBL: M33807; AAA23426.1; -;  
 DB EMBL: AF000202; AAC74087.1; -;  
 DB EMBL: D90737; AAA5769.1; -;  
 DB EMBL: D90738; AAA5779.1; -;  
 DB PIR: JV0087; JV0087.









```

1234 ATGTGTTGTGTGAGCTTTTAAATGAAATGATGAAAGACAGATATACCG----- 1294
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
349 --CysProLeuGluInAspPheuArgLeu-- ThrGluProValValProLysAspTrp 366
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1285 -----GNGTGGCACTTTCAGACATCT 1302
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 367 GluGlnGluCysGlnLeuAlaSer 474
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
RESULT 6
PVAL_KAT
ID PPAL_KAT STANDARD; PRT; 423 AA.
AC P20411;
DT 01-FEB-1991 (Rev. 17, Created)
GI 115 1341 (Rev. 17, last sequence update)
DI 01-NOV-1995 (Rev. 42, last annotation update)
DE Lysosomal acid phosphatase precursor (EC 4.1.3.2) (LAP).
GN ACP2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID:10116;
RN 11;
SEQUENCE FROM N.A.
MEDLINE 89450410; PubMed-27645916;
RA Himeño M., Fujita H., Nozuchi Y., Kono A., Kato K.;
RT "Isolation and sequencing of a cDNA clone encoding acid phosphatase
in rat liver lysosomes."
Biochem. Biophys. Res. Commun. 162:1044-1053(1989).
-| CATALYTIC ACTIVITY: An orthophosphoric monoester + H(2)O = an
alcohol + phosphate.
-| SUBCELLULAR LOCATION: Lysosomal.
-| SIMILARITY: BELONGS TO THE HISTIDINE ACID PHOSPHATASE FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

```

DR	BMR1:	MZ7894:	AAA40744_1:		
DR	PIR:	A33395:	A33395:		
DR	HSSP:	P20646:	IRPA:		
DR	InterPro:	IPR00560:	HISAc_phosphatase,		
DR	Fam:	PF00428:	acid phosphatase,		
DR	PROSITE:	PS00616:	HIS_ACID_PHOSPHAT_1:	1.	
DR	PROSITE:	PS00778:	HIS_ACID_PHOSPHAT_2:	1.	
DR	KW	Hydrolase:	SIGNAL: Glycoprotein; lysosome.		
FE	SIGNAL:	1	40		
FE	CHAIN:	41	423		LYSOSOMAL ACID PHOSPHATASE.
FE	DISEUFD:	159	370		BY SIMILARITY.
FE	DISEUFD:	212	310		BY SIMILARITY.
FE	DISEUFD:	345	349		BY SIMILARITY.
FE	ACT_SITE:	42	42		NUCLEOPHILIC ACCEPTOR (BY SIMILARITY).
FE	ACT_SITE:	84	84		BY SIMILARITY.
FE	ACT_SITE:	286	286		PROTON DONOR (BY SIMILARITY).
FE	CARBHYD:	92	92		N-LINKED (GLNAC... ) (POTENTIAL).
FE	CARBHYD:	143	143		N-LINKED (GLNAC... ) (POTENTIAL).
FE	CARBHYD:	147	167		N-LINKED (GLNAC... ) (POTENTIAL).
FE	CARBHYD:	177	177		N-LINKED (GLNAC... ) (POTENTIAL).
FE	CARBHYD:	191	191		N-LINKED (GLNAC... ) (POTENTIAL).
FE	CARBHYD:	197	197		N-LINKED (GLNAC... ) (POTENTIAL).
FE	CARBHYD:	267	267		N-LINKED (GLNAC... ) (POTENTIAL).
FE	CARBHYD:	322	322		N-LINKED (GLNAC... ) (POTENTIAL).
FE	CARBHYD:	331	331		N-LINKED (GLNAC... ) (POTENTIAL).
FE	SEQUENCE:	423 AA:	48319 MW:	6F4C4819CADD4496 CRC64;	
ASO	Alignment Scores:				
	Prod. No.:	0-0395	Length:	423	
	Score:	119.00	Matches:	105	
	%Identical Similarity:	92.90%	Conservative:	46	

[illegible]









type, maintaining the functional differentiation of such cells.

1- SUBCELLULAR LOCATION: Nuclear.

1- ALTERNATIVE PRODUCTS: 5 isoforms: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241, 242, 243, 244, 245, 246, 247, 248, 249, 250, 251, 252, 253, 254, 255, 256, 257, 258, 259, 260, 261, 262, 263, 264, 265, 266, 267, 268, 269, 270, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, 285, 286, 287, 288, 289, 290, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301, 302, 303, 304, 305, 306, 307, 308, 309, 310, 311, 312, 313, 314, 315, 316, 317, 318, 319, 320, 321, 322, 323, 324, 325, 326, 327, 328, 329, 330, 331, 332, 333, 334, 335, 336, 337, 338, 339, 340, 341, 342, 343, 344, 345, 346, 347, 348, 349, 350, 351, 352, 353, 354, 355, 356, 357, 358, 359, 360, 361, 362, 363, 364, 365, 366, 367, 368, 369, 370, 371, 372, 373, 374, 375, 376, 377, 378, 379, 380, 381, 382, 383, 384, 385, 386, 387, 388, 389, 390, 391, 392, 393, 394, 395, 396, 397, 398, 399, 400, 401, 402, 403, 404, 405, 406, 407, 408, 409, 410, 411, 412, 413, 414, 415, 416, 417, 418, 419, 420, 421, 422, 423, 424, 425, 426, 427, 428, 429, 430, 431, 432, 433, 434, 435, 436, 437, 438, 439, 440, 441, 442, 443, 444, 445, 446, 447, 448, 449, 450, 451, 452, 453, 454, 455, 456, 457, 458, 459, 460, 461, 462, 463, 464, 465, 466, 467, 468, 469, 470, 471, 472, 473, 474, 475, 476, 477, 478, 479, 480, 481, 482, 483, 484, 485, 486, 487, 488, 489, 490, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500, 501, 502, 503, 504, 505, 506, 507, 508, 509, 510, 511, 512, 513, 514, 515, 516, 517, 518, 519, 520, 521, 522, 523, 524, 525, 526, 527, 528, 529, 530, 531, 532, 533, 534, 535, 536, 537, 538, 539, 540, 541, 542, 543, 544, 545, 546, 547, 548, 549, 550, 551, 552, 553, 554, 555, 556, 557, 558, 559, 560, 561, 562, 563, 564, 565, 566, 567, 568, 569, 570, 571, 572, 573, 574, 575, 576, 577, 578, 579, 580, 581, 582, 583, 584, 585, 586, 587, 588, 589, 590, 591, 592, 593, 594, 595, 596, 597, 598, 599, 600, 601, 602, 603, 604, 605, 606, 607, 608, 609, 610, 611, 612, 613, 614, 615, 616, 617, 618, 619, 620, 621, 622, 623, 624, 625, 626, 627, 628, 629, 630, 631, 632, 633, 634, 635, 636, 637, 638, 639, 640, 641, 642, 643, 644, 645, 646, 647, 648, 649, 650, 651, 652, 653, 654, 655, 656, 657, 658, 659, 660, 661, 662, 663, 664, 665, 666, 667, 668, 669, 670, 671, 672, 673, 674, 675, 676, 677, 678, 679, 680, 681, 682, 683, 684, 685, 686, 687, 688, 689, 690, 691, 692, 693, 694, 695, 696, 697, 698, 699, 700, 701, 702, 703, 704, 705, 706, 707, 708, 709, 710, 711, 712, 713, 714, 715, 716, 717, 718, 719, 720, 721, 722, 723, 724, 725, 726, 727, 728, 729, 730, 731, 732, 733, 734, 735, 736, 737, 738, 739, 740, 741, 742, 743, 744, 745, 746, 747, 748, 749, 750, 751, 752, 753, 754, 755, 756, 757, 758, 759, 760, 761, 762, 763, 764, 765, 766, 767, 768, 769, 770, 771, 772, 773, 774, 775, 776, 777, 778, 779, 780, 781, 782, 783, 784, 785, 786, 787, 788, 789, 790, 791, 792, 793, 794, 795, 796, 797, 798, 799, 800, 801, 802, 803, 804, 805, 806, 807, 808, 809, 810, 811, 812, 813, 814, 815, 816, 817, 818, 819, 820, 821, 822, 823, 824, 825, 826, 827, 828, 829, 830, 831, 832, 833, 834, 835, 836, 837, 838, 839, 840, 841, 842, 843, 844, 845, 846, 847, 848, 849, 850, 851, 852, 853, 854, 855, 856, 857, 858, 859, 860, 861, 862, 863, 864, 865, 866, 867, 868, 869, 870, 871, 872, 873, 874, 875, 876, 877, 878, 879, 880, 881, 882, 883, 884, 885, 886, 887, 888, 889, 890, 891, 892, 893, 894, 895, 896, 897, 898, 899, 900, 901, 902, 903, 904, 905, 906, 907, 908, 909, 910, 911, 912, 913, 914, 915, 916, 917, 918, 919, 920, 921, 922, 923, 924, 925, 926, 927, 928, 929, 930, 931, 932, 933, 934, 935, 936, 937, 938, 939, 940, 941, 942, 943, 944, 945, 946, 947, 948, 949, 950, 951, 952, 953, 954, 955, 956, 957, 958, 959, 960, 961, 962, 963, 964, 965, 966, 967, 968, 969, 970, 971, 972, 973, 974, 975, 976, 977, 978, 979, 980, 981, 982, 983, 984, 985, 986, 987, 988, 989, 990, 991, 992, 993, 994, 995, 996, 997, 998, 999, 1000.

1- TISSUE SPECIFICITY: EXPRESSED IN THE EXCRETORY SYSTEM, THYROID GLAND AND WILMS TUMORS.

1- DEVELOPMENTAL STAGE: IN DEVELOPING EXCRETORY SYSTEM, DURING THYROID DIFFERENTIATION AND IN ADULT THYROID.

1- DISEASE: Defects in PAX8 are a cause of thyroid dysgenesis (TD), the most frequent cause of congenital hypothyroidism (CH). The thyroid gland can be absent (athyreosis), ectopically located (most often at the base of the tongue) and/or severely reduced in size (hypoplasia).

1- SIMILARITY: CONTAINS 1 PAIRED BOX DOMAIN.

This SWISS-Prot entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement. See <http://www.isb.sib.ch/autonomy> or send an email to [license@isb.sib.ch](mailto:license@isb.sib.ch).

EMBL: X69699; ; NOT\_ANNOTATED\_CDS.

EMBL: L19606; AAB04539.1; ;

EMBL: S77904; AAB34216.1; ;

EMBL: S77905; AAB34217.2; ;

EMBL: S77906; AAB34218.2; ;

PIR: A91263; A91263; ;

PIR: A91263; A91263; ;

PIR: A91263; A91263; ;

PIR: A91263; A91263; ;

PIR: A91263; A91263; ;

PIR: A91263; A91263; ;

PIR: A91263; A91263; ;

PIR: A91263; A91263; ;

PIR: A91263; A91263; ;

PIR: A91263; A91263; ;

PIR: A91263; A91263; ;

PIR: A91263; A91263; ;

PIR: A91263; A91263; ;

PIR: A91263; A91263; ;

PIR: A91263; A91263; ;

PIR: A91263; A91263; ;

PIR: A91263; A91263; ;

PIR: A91263; A91263; ;

PIR: A91263; A91263; ;

PIR: A91263; A91263; ;

PIR: A91263; A91263; ;

PIR: A91263; A91263; ;

PIR: A91263; A91263; ;

PIR: A91263; A91263; ;

PIR: A91263; A91263; ;

PIR: A91263; A91263; ;

PIR: A91263; A91263; ;

PIR: A91263; A91263; ;

PIR: A91263; A91263; ;

PIR: A91263; A91263; ;

PIR: A91263; A91263; ;

PIR: A91263; A91263; ;

PIR: A91263; A91263; ;

PIR: A91263; A91263; ;

PIR: A91263; A91263; ;

PIR: A91263; A91263; ;

PIR: A91263; A91263; ;

PIR: A91263; A91263; ;

PIR: A91263; A91263; ;

PIR: A91263; A91263; ;

PIR: A91263; A91263; ;

PIR: A91263; A91263; ;

PIR: A91263; A91263; ;

PIR: A91263; A91263; ;

PIR: A91263; A91263; ;

PIR: A91263; A91263; ;

PIR: A91263; A91263; ;

PIR: A91263; A91263; ;

PIR: A91263; A91263; ;

PIR: A91263; A91263; ;

PIR: A91263; A91263; ;

PIR: A91263; A91263; ;

PIR: A91263; A91263; ;

PIR: A91263; A91263; ;

PIR: A91263; A91263; ;

PIR: A91263; A91263; ;









```

DR InterPro: IPR004358; Bact_sens_pr_C.
DR InterPro: IPR003661; His_kinA.
DR InterPro: IPR004359; HIS_KIN_sig.
DR InterPro: IPR000014; PAS_domain.
DR Pfam: PF02518; HATPase_C_1.
DR Pfam: PF00512; signal_1.
DR PRINTS: PR00344; RCIRUSENSOR.
DR SMART: SM00387; HATPase_C_1.
DR SMART: SM00388; HisKA_1.
DR TIGPAMS: TIGP00229; sensory_box_2.
DR PROSITE: PS0109; HIS_KIN_1.
KW Sensory transduction; Transferase; Kinase; Transmembrane;
KW Phosphorylation; Complete proteome.
FT DOMAIN 6 26 POTENTIAL..
FT DOMAIN 547 758 HISTIDINE KINASE.
FT DOMAIN 9 221 ALA-RICH.
FT MOTIF_RES 550 550 PHOSPHORYLATION (AUTO-).
FT CONFLICT 200 200 Q -> H (IN REF. 1).
FT CONFLICT 216 216 V -> E (IN REF. 1).
SQ SEQUENCE 769 AA; 82796 MW; 002B2428F18A57EF CRC64;

Alignment Scores:
Pred. No.: 0.338 Length: 769
Score: 108.50 Matches: 102
Percent Similarity: 31.94% Conservative: 36
Best Local Similarity: 23.61% Mismatches: 147
Query Match: 4.70% Indels: 148
Gaps: 20

US-09-866-479A-9 (1-1308) x DIVL_CAOCR (1-769)
QY 136 AAGGACAGCAACTG-----ATCAGCATGTGCACGACAGCATGCGCAACCTGGCG 189
DB 135 AAGGACAGCAACTG-----ATCAGCATGTGCACGACAGCATGCGCAACCTGGCG 189
DB 135 AAGGACAGCAACTG-----ATCAGCATGTGCACGACAGCATGCGCAACCTGGCG 189
QY 345 LysAlaLueGluLeuAlaAra:47rGluAspLeuGlyProGluAlaAspLeuTrpAsp 364
DB 345 LysAlaLueGluLeuAlaAra:47rGluAspLeuGlyProGluAlaAspLeuTrpAsp 364
QY 190 GTAAACTGGGTGAGCTGACA-----CCGCGGGGGTGGTGGAGCTA 228
DB 190 GTAAACTGGGTGAGCTGACA-----CCGCGGGGGTGGTGGAGCTA 228
DB 190 GTAAACTGGGTGAGCTGACA-----CCGCGGGGGTGGTGGAGCTA 228
QY 365 LeuProAspGlyAcqThsLeuLysValValArqGlnProHisProLeuGlyGlyMetLeu 384
DB 365 LeuProAspGlyAcqThsLeuLysValValArqGlnProHisProLeuGlyGlyMetLeu 384
QY 229 AWCAGCTATCTGGACATATGCGGT----- 255
DB 229 AWCAGCTATCTGGACATATGCGGT----- 255
DB 229 AWCAGCTATCTGGACATATGCGGT----- 255
QY 385 LeuLeuTyrSerAspIleThrGlyGluLeuArqLeuLysAlaGlnTrpAsnAlaLeuIle 404
DB 385 LeuLeuTyrSerAspIleThrGlyGluLeuArqLeuLysAlaGlnTrpAsnAlaLeuIle 404
QY 256 -----CAAGCTCTGTAGCGACGATTTCTGTGCTAAATCTGGTCTGCGGACGCTGG 308
DB 256 -----CAAGCTCTGTAGCGACGATTTCTGTGCTAAATCTGGTCTGCGGACGCTGG 308
DB 256 -----CAAGCTCTGTAGCGACGATTTCTGTGCTAAATCTGGTCTGCGGACGCTGG 308
QY 404 cclnValGlnGlnAlaThrLeuAspLysLeu-----AsnAs 416
DB 404 cclnValGlnGlnAlaThrLeuAspLysLeu-----AsnAs 416
QY 409 TCAGCTGGCGATATTGCTGATCTGACGAGGCTAAGGTAAGGTAAGGTAAGGTAAGGTAAG 368
DB 409 TCAGCTGGCGATATTGCTGATCTGACGAGGCTAAGGTAAGGTAAGGTAAGGTAAGGTAAG 368
DB 409 TCAGCTGGCGATATTGCTGATCTGACGAGGCTAAGGTAAGGTAAGGTAAGGTAAGGTAAG 368
QY 416 pAlaValAlaValPheGlySer---AspGlyArqLeuArqLeuHisAsnGluAlaPheG 435
DB 416 pAlaValAlaValPheGlySer---AspGlyArqLeuArqLeuHisAsnGluAlaPheG 435
QY 469 CGCC-----GGGCGGCG 380
DB 469 CGCC-----GGGCGGCG 380
DB 469 CGCC-----GGGCGGCG 380
QY 435 aThrPheTrpAsnValThrPheHisAlaLeuGluAlaLysGlyAspPheGluGlyValVa 455
DB 435 aThrPheTrpAsnValThrPheHisAlaLeuGluAlaLysGlyAspPheGluGlyValVa 455
QY 381 AWCAGCTGTGCAATAACGCGATAT-----ACCGAGGACAGATAC 419
DB 381 AWCAGCTGTGCAATAACGCGATAT-----ACCGAGGACAGATAC 419
DB 381 AWCAGCTGTGCAATAACGCGATAT-----ACCGAGGACAGATAC 419
QY 455 lclnLeuGlyValProArqGluHisAspLeuSerPheTrpA:ggLueLysGlyArgVa 475
DB 455 lclnLeuGlyValProArqGluHisAspLeuSerPheTrpA:ggLueLysGlyArgVa 475
QY 420 GTCCACTCGGATCGGTATT----- 441
DB 420 GTCCACTCGGATCGGTATT----- 441
DB 420 GTCCACTCGGATCGGTATT----- 441
QY 475 lAlaAspLeuAspLeuGlnMetAlaAlaProGluSerGlyGluValArqThrSerAspSe 495
DB 475 lAlaAspLeuAspLeuGlnMetAlaAlaProGluSerGlyGluValArqThrSerAspSe 495
QY 442 -----AATGCTCTATAAAGTGGGTTGGCAATTCATACATACACCG 479
DB 442 -----AATGCTCTATAAAGTGGGTTGGCAATTCATACATACACCG 479
DB 442 -----AATGCTCTATAAAGTGGGTTGGCAATTCATACATACACCG 479
QY 495 rArqIleValLeuTyrGlnSerArqProLeuProAspGlyAlaThrLeuIleAlaPheAl 515
DB 495 rArqIleValLeuTyrGlnSerArqProLeuProAspGlyAlaThrLeuIleAlaPheAl 515
QY 480 GAAGCTGATTCAGGATCTGTCAGAGGACACAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 539
DB 480 GAAGCTGATTCAGGATCTGTCAGAGGACACAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 539
DB 480 GAAGCTGATTCAGGATCTGTCAGAGGACACAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 539
QY 515 aAspValThrAspThr----- 520
DB 515 aAspValThrAspThr----- 520
QY 540 TTATCAAAAGCGGTTCTCGCAAACTGCAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 569
DB 540 TTATCAAAAGCGGTTCTCGCAAACTGCAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 569
DB 540 TTATCAAAAGCGGTTCTCGCAAACTGCAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 569

```

```

Db 521 -----ArqAspLeuGlnSerAlaLeu----- 527
QY 600 GTTTAAACCTCAGAAACAGCAGAAAGCTGTTTCATTAAAGAGGATTAACATATGAACT 659
DB 600 GTTTAAACCTCAGAAACAGCAGAAAGCTGTTTCATTAAAGAGGATTAACATATGAACT 659
DB 600 GTTTAAACCTCAGAAACAGCAGAAAGCTGTTTCATTAAAGAGGATTAACATATGAACT 659
Db 528 -----AlaAspArgSerAlaAlaLeuAlaGluAla-----G 538
QY 660 AAGGATTAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 713
DB 660 AAGGATTAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 713
DB 660 AAGGATTAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 713
Db 538 uArgLeuLysArgAspPheVal-----GlyAsnValSerTyrGlnLeuAlaValThr 555
QY 714 GTCAGCGCAGATATT-----CTCTCTCAACCAAGCAGCAGCAGCAGCAGCAGCAGCAG 761
DB 714 GTCAGCGCAGATATT-----CTCTCTCAACCAAGCAGCAGCAGCAGCAGCAGCAGCAG 761
DB 714 GTCAGCGCAGATATT-----CTCTCTCAACCAAGCAGCAGCAGCAGCAGCAGCAGCAG 761
Db 555 oLeuThrThrIleIleGlyTyrSerGluLeuGluAlaAlaAspGlyIleSerGluAla 575
QY 762 GTCAGCGCAGATATT-----CTCTCTCAACCAAGCAGCAGCAGCAGCAGCAGCAGCAG 810
DB 762 GTCAGCGCAGATATT-----CTCTCTCAACCAAGCAGCAGCAGCAGCAGCAGCAGCAG 810
DB 762 GTCAGCGCAGATATT-----CTCTCTCAACCAAGCAGCAGCAGCAGCAGCAGCAGCAG 810
Db 575 gGlyArgAsnHisValAlaAlaValArqAlaAlaThrGlnLeuAlaArqSerIleAs 595
QY 811 -----TTGCATATAGCGCAATTGATTGCTACAGGAGAGGAGGAGGAGGAGGAGGAG 863
DB 811 -----TTGCATATAGCGCAATTGATTGCTACAGGAGAGGAGGAGGAGGAGGAGGAG 863
DB 811 -----TTGCATATAGCGCAATTGATTGCTACAGGAGAGGAGGAGGAGGAGGAGGAG 863
Db 595 pAspValLeuAspMetAlaGlnIleAspAlaGlyGluMetAlaLeuGlnIleAsp 615
QY 864 GTCAGCGCAGATATT-----CTCTCTCAACCAAGCAGCAGCAGCAGCAGCAGCAGCAG 923
DB 864 GTCAGCGCAGATATT-----CTCTCTCAACCAAGCAGCAGCAGCAGCAGCAGCAGCAG 923
DB 864 GTCAGCGCAGATATT-----CTCTCTCAACCAAGCAGCAGCAGCAGCAGCAGCAGCAG 923
Db 615 eArgValSer-----AspLeuLeuLeuAsnAlaGlnIleAlaAlaLeuLysAspAl 632
QY 924 ACAGCGGATATGGTGTGACATACACCATTCATCTCTGTTATCTGAGGAGGAGGAGGAG 983
DB 924 ACAGCGGATATGGTGTGACATACACCATTCATCTCTGTTATCTGAGGAGGAGGAGGAG 983
DB 924 ACAGCGGATATGGTGTGACATACACCATTCATCTCTGTTATCTGAGGAGGAGGAGGAG 983
Db 632 acIleLeuGlyGlyValThrLeuAlaValGlu-----CysGlnGluAspValG 648
QY 984 TGTGCGCAATCTGCGCGCGCGCATGCACTCACTCACTCACTCACTCACTCACTCACTCACT 1043
DB 984 TGTGCGCAATCTGCGCGCGCGCATGCACTCACTCACTCACTCACTCACTCACTCACTCACT 1043
DB 984 TGTGCGCAATCTGCGCGCGCGCATGCACTCACTCACTCACTCACTCACTCACTCACTCACT 1043
Db 648 rLeuIleAlaGlyAspGlyGlyValThrLeuAlaValGlu-----CysGlnGluAspValG 667
QY 1044 C-----ACGCGCGCGCGCGCGCATGCACTCACTCACTCACTCACTCACTCACTCACT 1089
DB 1044 C-----ACGCGCGCGCGCGCGCATGCACTCACTCACTCACTCACTCACTCACTCACT 1089
DB 1044 C-----ACGCGCGCGCGCGCGCATGCACTCACTCACTCACTCACTCACTCACTCACT 1089
Db 667 nAlaLeuArgGlnThrProGlyGlyValThrLeuSerAlaAlaArqAlaLeuGlu 687
QY 1090 -----AGCGATACACGCGCGCGCATGCACTCACTCACTCACTCACTCACTCACTCACT 1130
DB 1090 -----AGCGATACACGCGCGCGCATGCACTCACTCACTCACTCACTCACTCACTCACT 1130
DB 1090 -----AGCGATACACGCGCGCGCATGCACTCACTCACTCACTCACTCACTCACTCACT 1130
Db 687 yGluValArqLeuAspValSerAsp-ThrGlyArgGlyValProPheHisValGlnAla 707
QY 1131 GACTTTTACAGCATGCGTGCATATAAACGCGCGC 1162
DB 1131 GACTTTTACAGCATGCGTGCATATAAACGCGCGC 1162
DB 1131 GACTTTTACAGCATGCGTGCATATAAACGCGCGC 1162
Db 707 isIlePheAspArgPheValGlyArgAspArg 717
QY 1131 GACTTTTACAGCATGCGTGCATATAAACGCGCGC 1162
DB 1131 GACTTTTACAGCATGCGTGCATATAAACGCGCGC 1162
DB 1131 GACTTTTACAGCATGCGTGCATATAAACGCGCGC 1162
RESULT 14
PPAP_HUMAN STANDARD; PRT; 386 AA.
AC P15309;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Prostatic acid phosphatase precursor (EC 3.1.3.2).
GN ACPP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92572747; PubMed=1375464;
RA Sharief F.S.; Li S.S.-L.;
RT "Structure of human prostatic acid phosphatase gene.";
RL Biochem. Biophys. Res. Commun 184:1468-1476(1992).
RN [2]
RP SEQUENCE FROM N.A., PARTIAL SEQUENCE, DISULFIDE BONDS, AND ACTIVE SITE.
RX MEDLINE=91115848; PubMed=1989985;
RA van Eften K.L.; Davidson F.; Stevis P.F.; MacArthur H.; Moore D.L.;
RT "Covalent structure, disulfide bonding, and identification of reactive surface and active site residues of human prostatic acid phosphatase.";

```

J. Biol. Chem. 266:2313-2319(1991).  
[4] SEQUENCE FROM N.A.  
MEDLINE 89228054; PubMed 2712844.  
RA Sharlet F.S., Lee H., Leidenman M.M., Lundwall A., Deaven L.L.,  
RA Lee C.-L., Li S.S.-L.;  
RT "Human prostatic acid phosphatase: cDNA cloning, gene mapping and  
RT protein sequence homology with lysosomal acid phosphatase.";  
RT Biochem. Biophys. Res. Commun. 160:79-86(1989).  
[4] SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RT TISSUE-Prostate;  
RC MF01NF-BR11981; PubMed-2942194;  
RA Vihko P., Virkkunen P., Heitto P., Reiko K., Salla T., Huhtala M.L.,  
RT "Molecular cloning and sequence analysis of cDNA encoding human  
RT prostatic acid phosphatase.";  
KL FEBS Lett. 236:275-281(1988).  
[5] SEQUENCE FROM N.A.  
RT TISSUE-Prostate;  
RC MEDLINE 90370491; PubMed-2395659;  
RA Taylor P.G., Govindan M.V., Patel P.C.;  
RT "Nucleotide sequence of human prostatic acid phosphatase determined  
RT from a full-length cDNA clone.";  
PL Nucleic Acids Res. 18:4928-4928(1990).  
[6] SEQUENCE FROM N.A.  
RT MEDLINE 95038546; PubMed 7951074;  
RA Sharlet F.S., Li S.S.-L.;  
RT "Nucleotide sequence of human prostatic acid phosphatase ACPP gene,  
RT including seven Alu repeats.";  
KL Biochem. Mol. Biol. Int. 33:561-565(1994).  
[7] X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).  
RT MEDLINE 99023966; PubMed 9804809;  
RA Lacomit M.W., Handy G., Lebioda L.;  
RT "Structural origins of L-lysine inhibition of human prostatic  
RT acid phosphatase.";  
KL J. Biol. Chem. 273:30406-30409(1998).  
[8] CATALYTIC ACTIVITY: An orthophosphoric monoester + H(2)O = an  
RT alcohol + phosphate.  
[8] SUBUNIT: HOMODIMER.  
[8] SIMILARITY: BELONGS TO THE HISTIDINE ACID PHOSPHATASE FAMILY.  
[8] THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC at the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb.sib.ch/announc>  
CC or send an email to [license@isb.sib.ch](mailto:license@isb.sib.ch)).

QY	7	GGAATCTAAATGACATTTTATTCTCTGGCATTCGTAAAGCCCAAAATGCATCGT	66
Dd			
	13	SerLeuSerLeuGlyPheIeuPhoIeuLeuPhoIeuPheIeuPheIeuAspArgSerValIleuAla	42
QY	97	CAGATTCAGAGGAGATGAAAGATGAAAAAGTGAGTCAATTAACAATGATAATGAT	129
Dd			
Dd	33	Lys-----GluIeuLysPheValThrLeuValPheArgHisSerIleAspArg	47
QY	127	GGTCCAAAAGAAGGCAAGCAACTGATGAGCAGATGAGCAAGCAAGCAAGCAAGCAAG	183
Dd			
	48	SerProIle-----AspThrPheThrThrAsp	69
QY	184	-----TGSGTAGCTAATAAACCTGGGTGGAGCTCAAGCCAGCCAGCCAGCCAGCCAG	225
Dd			
	57	ProIleLysGluSerSerTrpMetIndIlyPheGlyIleuIeuIleuIndIlyGlyMetThr	99
QY	226	CTAATCGCGTATCTGCCAATTACTGAGAGCAAGCT	276
Dd			
	77	GlnHisTryIleuIeuLysIleuTyrThrAlaGlySerArgArgAlaLysPheLeuAlaHisGlu	99
QY	277	TTCGTGCTAAATGTGATGTGTGGAGTCTGGTAACTGGTAACTGGTAACTGGTAACTGGT	309





[illegible][illegible]

Search completed: December 14, 2002, 16:03:08  
Job time : 46.5 secs





C:Genetics:

A:Gene: EC51146

Alignment Scores:

Pred. No.: 2,82e-170 Length: 434  
 Score: 2144.00 Matches: 419  
 Percent Similarity: 96.54% Conservativeness: 0  
 Best Local Similarity: 96.54% Mismatches: 13  
 Query Match: 92.85% Indels: 2  
 DB: 2 Gaps: 1

US-09-866-379A-9 (1-1308) x H90770 (1-434)

QY 1 ATGAAGGAGTCTTATCCCAATTTTATCTCTTCGATCCGTTAACCCGCCAATCTGCA 60  
 DB 1 MetLysAlaIleLeuLeuProPheLeuSerLeuLeuLeuProLeuThrProGlnSerAla 20  
 QY 61 TTTCCTCAGAGT-----GAGCCGAGCTGAAGCTGCAAGCTGCTGCTCAGTCTGCT 114  
 DB 21 PheAlaGlnSerGluProGluProGluLeuLysLeuGluSerValIleValSerArg 40  
 QY 115 CATGCTGTGTGTGTAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 174  
 DB 41 HisGlyValArgAlaLeuThrLysAlaThrGlnLeuMetGlnAspValThrProAspAla 60  
 QY 175 TGCGCAACCTGGCGGCTAAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 234  
 DB 61 TrpProAsnTrpProValLysLeuGlyTrpLeuThrProArgGlyGlyGlnLeuIleAla 80  
 QY 235 TATCTCGCAATTAATGGGCGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 294  
 DB 81 TyrLeuGlyHisTyrGlnArgGlnArgLeuValAlaAspGlyLeuLeuThrLysLysGly 100  
 QY 295 TGGGCGAGT 354  
 DB 101 CysProGlnProClyGlnValAlaIleLeuAlaAspValAspClyGlnValThrArgLysThr 120  
 QY 355 GCGCAAGCTGT 414  
 DB 121 GlyLeuAlaProAlaAlaGlyLeuAlaProAspGlyAlaIleThrValHisThrGlnAla 140  
 QY 415 GATAGTCCAGTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 474  
 DB 141 AspThrSerSerProAspProLeuPheAsnTrpLeuLysThrGlyValCysGlnLeuAsp 160  
 QY 475 AAGCGGAAGT 534  
 DB 161 AsnAlaAsnValThrAspAlaIleLeuSerArgAlaGlySerIleAlaAspPheThr 180  
 QY 535 GAGCATTAATGAAGCGGCTTTGGCGAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 594  
 DB 181 GlyHisArgLeuThrAlaGlyAlaArgGlyGlnArgGlyValLeuAsnProGlnSerAsn 600  
 QY 595 TTGTCGCTTAAACCGGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGAC 654  
 DB 201 LeuCysLeuAsnArgGlyGlnAspGlnSerCysSerLeuThrGlnAlaLeuProSer 220  
 QY 655 GAACCTAAGT 714  
 DB 221 GluLeuGlyValSerAlaAspAsnValSerLeuThrGlyAlaValSerLeuAlaSerMet 240  
 QY 715 GAGAGGAGAGATTTCTGT 774  
 DB 241 LeuThrGlyIlePheLeuLeuValAlaIleLeuValMetProGluProGlyTrpGlyArg 260  
 QY 775 ATCAGCGATTACACAGT 834  
 DB 261 IleThrAspSerHisGlnTrpAsnThrLeuLeuSerLeuHisAsnAlaGlnPheThrLeu 280  
 QY 835 GTATCAAGT 894  
 DB 281 LeuGlnArgThrProGluValAlaArgSerAlaAlaThrProLeuLeuAspLeuIleMet 300

QY 895 ACAGCGTTACGCGCCCATCCACGCAAAACAGCGGTATGAGTGTGACATTAACGATTTCA 954  
 DB 301 IleAlaLeuThrProHisProGlnLysGlnAlaTyrGlyValThrLeuProThrSer 420  
 QY 955 GT 1014  
 DB 321 ValLeuProGlyAlaGlyHisAspThrAsnLeuAlaLeuLeuGlyGlyAlaLeuLeuLeu 340  
 QY 1015 AACTGAGCGCTTCCCGGTCAACGCGGATAACAGCGGCGCGAGGTGTGTGTGTGTGTGT 1074  
 DB 341 AsnTrpThrLeuProGlyClyLeuProAspAsnThrProProGlyGlyClyLeuValPheGlu 460  
 QY 1075 GCTTGGCTGCGCTTAAGCGATAACAGCGAGTGGATTCAGTTCCTGCTGCTTTCACAT 1134  
 DB 361 ArgTrpArgArgLeuSerAspAsnSerGlnTrpIleGlnValSerLeuValPheGlnThr 480  
 QY 1135 TTACAGCAGATCGCTGTGATAAAGCGCGCTGTTCATTAAATACGCGCGGTGTGTGTGT 1194  
 DB 381 LeuGlnGlnMetArgAspLysThrProLeuSerLeuAsnThrProProGlyGluValLys 400  
 QY 1195 GAGACCTGT 1254  
 DB 401 LeuThrLeuAlaGlyCysGluArgAsnAlaGlyGlyMetCysSerLeuAlaGlyPhe 420  
 QY 1255 ACGCAAAATCGTAATGAAGCAAGCATACCGCGCTGTGACTTTG 1296  
 DB 421 ThrGlnIleValAsnGluAlaArgIleProAlaCysSerLeu 434

## RESULT 3

p85633

hypothetical protein appA [imported] - Escherichia coli (strain 0157:H7, substrain  
 C:Species: Escherichia coli  
 C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #ext\_change 14 Sep-2001  
 C:Accession: D85633

R:Perla, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.;

Nelson, K.E.; Burland, V.; Davis, N.W.; Tim, A.; Dimalanta, E.; Potamousis, K.;

Nelson, K.E.; Burland, V.; Davis, N.W.; Tim, A.; Dimalanta, E.; Potamousis, K.;

Nelson, K.E.; Burland, V.; Davis, N.W.; Tim, A.; Dimalanta, E.; Potamousis, K.;

Nelson, K.E.; Burland, V.; Davis, N.W.; Tim, A.; Dimalanta, E.; Potamousis, K.;

Nelson, K.E.; Burland, V.; Davis, N.W.; Tim, A.; Dimalanta, E.; Potamousis, K.;

Nelson, K.E.; Burland, V.; Davis, N.W.; Tim, A.; Dimalanta, E.; Potamousis, K.;

Nelson, K.E.; Burland, V.; Davis, N.W.; Tim, A.; Dimalanta, E.; Potamousis, K.;

Nelson, K.E.; Burland, V.; Davis, N.W.; Tim, A.; Dimalanta, E.; Potamousis, K.;

Nelson, K.E.; Burland, V.; Davis, N.W.; Tim, A.; Dimalanta, E.; Potamousis, K.;

Nelson, K.E.; Burland, V.; Davis, N.W.; Tim, A.; Dimalanta, E.; Potamousis, K.;

Nelson, K.E.; Burland, V.; Davis, N.W.; Tim, A.; Dimalanta, E.; Potamousis, K.;

Nelson, K.E.; Burland, V.; Davis, N.W.; Tim, A.; Dimalanta, E.; Potamousis, K.;

Nelson, K.E.; Burland, V.; Davis, N.W.; Tim, A.; Dimalanta, E.; Potamousis, K.;

Nelson, K.E.; Burland, V.; Davis, N.W.; Tim, A.; Dimalanta, E.; Potamousis, K.;

Nelson, K.E.; Burland, V.; Davis, N.W.; Tim, A.; Dimalanta, E.; Potamousis, K.;

Nelson, K.E.; Burland, V.; Davis, N.W.; Tim, A.; Dimalanta, E.; Potamousis, K.;

Nelson, K.E.; Burland, V.; Davis, N.W.; Tim, A.; Dimalanta, E.; Potamousis, K.;

Nelson, K.E.; Burland, V.; Davis, N.W.; Tim, A.; Dimalanta, E.; Potamousis, K.;

Nelson, K.E.; Burland, V.; Davis, N.W.; Tim, A.; Dimalanta, E.; Potamousis, K.;

Nelson, K.E.; Burland, V.; Davis, N.W.; Tim, A.; Dimalanta, E.; Potamousis, K.;

Nelson, K.E.; Burland, V.; Davis, N.W.; Tim, A.; Dimalanta, E.; Potamousis, K.;

Nelson, K.E.; Burland, V.; Davis, N.W.; Tim, A.; Dimalanta, E.; Potamousis, K.;

Nelson, K.E.; Burland, V.; Davis, N.W.; Tim, A.; Dimalanta, E.; Potamousis, K.;

Nelson, K.E.; Burland, V.; Davis, N.W.; Tim, A.; Dimalanta, E.; Potamousis, K.;

Nelson, K.E.; Burland, V.; Davis, N.W.; Tim, A.; Dimalanta, E.; Potamousis, K.;

Nelson, K.E.; Burland, V.; Davis, N.W.; Tim, A.; Dimalanta, E.; Potamousis, K.;

Nelson, K.E.; Burland, V.; Davis, N.W.; Tim, A.; Dimalanta, E.; Potamousis, K.;

Nelson, K.E.; Burland, V.; Davis, N.W.; Tim, A.; Dimalanta, E.; Potamousis, K.;

Nelson, K.E.; Burland, V.; Davis, N.W.; Tim, A.; Dimalanta, E.; Potamousis, K.;

Nelson, K.E.; Burland, V.; Davis, N.W.; Tim, A.; Dimalanta, E.; Potamousis, K.;

Nelson, K.E.; Burland, V.; Davis, N.W.; Tim, A.; Dimalanta, E.; Potamousis, K.;

Nelson, K.E.; Burland, V.; Davis, N.W.; Tim, A.; Dimalanta, E.; Potamousis, K.;

Nelson, K.E.; Burland, V.; Davis, N.W.; Tim, A.; Dimalanta, E.; Potamousis, K.;

Nelson, K.E.; Burland, V.; Davis, N.W.; Tim, A.; Dimalanta, E.; Potamousis, K.;

Nelson, K.E.; Burland, V.; Davis, N.W.; Tim, A.; Dimalanta, E.; Potamousis, K.;

Nelson, K.E.; Burland, V.; Davis, N.W.; Tim, A.; Dimalanta, E.; Potamousis, K.;

Nelson, K.E.; Burland, V.; Davis, N.W.; Tim, A.; Dimalanta, E.; Potamousis, K.;

Nelson, K.E.; Burland, V.; Davis, N.W.; Tim, A.; Dimalanta, E.; Potamousis, K.;

Nelson, K.E.; Burland, V.; Davis, N.W.; Tim, A.; Dimalanta, E.; Potamousis, K.;

Nelson, K.E.; Burland, V.; Davis, N.W.; Tim, A.; Dimalanta, E.; Potamousis, K.;

Nelson, K.E.; Burland, V.; Davis, N.W.; Tim, A.; Dimalanta, E.; Potamousis, K.;

Nelson, K.E.; Burland, V.; Davis, N.W.; Tim, A.; Dimalanta, E.; Potamousis, K.;

Nelson, K.E.; Burland, V.; Davis, N.W.; Tim, A.; Dimalanta, E.; Potamousis, K.;

Nelson, K.E.; Burland, V.; Davis, N.W.; Tim, A.; Dimalanta, E.; Potamousis, K.;

Nelson, K.E.; Burland, V.; Davis, N.W.; Tim, A.; Dimalanta, E.; Potamousis, K.;

Nelson, K.E.; Burland, V.; Davis, N.W.; Tim, A.; Dimalanta, E.; Potamousis, K.;

Nelson, K.E.; Burland, V.; Davis, N.W.; Tim, A.; Dimalanta, E.; Potamousis, K.;

Alignment Scores:

Pred. No.: 2,82e-170 Length: 444  
 Score: 2144.00 Matches: 419  
 Percent Similarity: 96.54% Conservativeness: 0  
 Best Local Similarity: 96.54% Mismatches: 13  
 Query Match: 92.85% Indels: 2  
 DB: 2 Gaps: 1

US-09-866-379A-9 (1-1308) x D85633 (1-444)

QY 1 ATGAAGGAGTCTTATCCCAATTTTATCTCTTCGATCCGTTAACCGGAAATCTGCA 60  
 DB 11 MetLysAlaIleLeuLeuProPheLeuSerLeuLeuLeuProLeuThrProGlnSerAla 40  
 QY 61 TTTCCTCAGAGT-----GAGCCGAGCTGAAGCTGCTGCTGCTGCTGCTGCTGCTGCT 114  
 DB 31 PheAlaGlnSerGluProGluProGluLeuLysLeuGluSerValIleValSerArg 50  
 QY 115 CATGCTGTGTGTGTAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 174  
 DB 51 HisGlyValArgAlaProThrLysAlaThrGlnLeuMetGlnAspValThrProAspAla 70  
 QY 175 TGCGCAACCTGGCGGCTAAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 234  
 DB 61 TrpProAsnTrpProValLysLeuGlyTrpLeuThrProArgGlyGlyGlnLeuIleAla 90

## RESULT 4

AC0201  
acid phosphatase (EC 3.1.3.2) [imported] Yersinia pestis (strain CO92)  
C:Species: Yersinia pestis  
C:Date: 02 Nov 2001 #sequence\_revision 92 Nov 2001 #size change 27 Nov 2001  
C:Accession: AC0201  
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Tibball, R.W.; Borden, M.; East, P.; Brudenell  
deno-Tarrava, A.M.; Chillingworth, J.; Cronin, A.; Davies, K.M.; Davies, P.; Dougan  
11, M.; Rutherford, K.; Simmonds, M.; Skellern, J.; Stevens, K.; Whitehead, P.; Barr  
Nature 413, 523-527, 2001  
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.  
A:Reference number: AB0001; MIMD:2147041; IMH01158660  
A:Accession: AC0201  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-44 ->KOR  
A:Cross-references: GB:AI590842; FIDN:CA790470.1; FIDN:CA790470.2; GB:DB:000000  
C:Genetics:  
A:Gene: Ypo1648  
C:Keywords: phosphatase, mononuclear hydrolase

Alignment Scores  
Pred. No.: 1,126-71 441  
Score: 959,50  
Percent Similarity: 61.47%  
Best Local Similarity: 46.10%  
Query Match: 41.55%  
Indels: 35  
Gaps: 2

US-09-866-479A-9 (1-1308) X AC0201 (1-441)

QY 13 TAAATACCATTTTATCTCTTCTGATCTGCTTAAATCTGCTTAAATCTGCT 294  
DB 13 LeuValLeuMetLeuSerGlyLeuAlaAlaThrAlaProValAla 294  
QY 74 GAGGCGAG--CGCAAG--TGGAAAGTGGTGGTGGTGGTGGTGGTGGTGGT 129  
DB 74 GluProSerGlyTyrThrLeuGluArgValValLeuLeuSerAlaLeuMetGlyGlyPhe 129  
QY 140 GAAATTAAGGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 189  
DB 140 GAAATTAAGGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 189  
QY 50 ProThrLysGlnThrGlnLeuMetAspAspValThrProAspLysThrProLeuThrPro 69  
DB 50 GTAAATCTGGCTGAGCTGCAATGCGGTGGTGGTGGTGGTGGTGGTGGTGGT 149  
DB 70 ValLysAlaLeuTyrLeuThrProArgValAlaLeuLeuValLeuLeuMetGlyGlyPhe 69  
QY 250 TGGCTGACGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 309  
DB 250 TGGCTGACGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 309  
QY 90 TyrGlyAspTyrPheAspSerLeuGlyLeuLeu AlaAlaLysGlySerAlaLeuThrGly 106  
QY 310 CAGGTGACGATTATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATG 369  
DB 109 GlyValTyrAlaGlnAlaAspLeuAspLeuAlaThrAlaLeuThrGlnAlaAlaLeu 128  
QY 370 GCGGAGTGGGACCGACGACGACGACGACGACGACGACGACGACGACGACG 429  
DB 129 AspLysValAlaProLysGlyTyrThrValThrLeuLeuLeuLeuLeuLeuLeu 146  
QY 430 GATCGATGATTAATGCTTAAATGCTTAAATGCTTAAATGCTTAAATGCTTAA 489  
DB 149 AspThrLeuPheHisProValGlnAlaGlyValCysLysLeuAspAlaAlaThrAsp 168  
QY 490 GAGCGATGCTGACGACGACGACGACGACGACGACGACGACGACGACGAC 549  
DB 169 LysAlaLeuLeuGlnGlnLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 188  
QY 550 GCGTTGCGCAACGCAACGCAACGCAACGCAACGCAACGCAACGCAACGCAAC 609  
DB 189 ProPheAlaLeuMetGlyAspValLeuAsnThrAlaAlaSerProLysCysSerLeu 208  
QY 610 GATCAAAAGACGAAAGCTGCTTCAATTAAGCTGCTTCAATTAAGCTGCTTCA 669  
DB 610 GATCAAAAGACGAAAGCTGCTTCAATTAAGCTGCTTCAATTAAGCTGCTTCA 669

Db 229 GluGlnIleuValGlySerThrValAspPheAlaValAsnGlnValAsnValAsn 228  
 QY 678 GGGAC-----TCCTCTCATACCGGTGGGTAAGCGCGGCGATCAATCTGACGAG 723  
 Db 229 GlyGlnValValThrLeuSerGlyProAlaValAsnSerThrLeuValGln 248  
 QY 724 ATATTCTCTGCAAC 783  
 Db 249 ThrLeuLeuGlnValAlaValAlaValAlaValAlaValAlaValAlaValAla 268  
 QY 784 TCAC 843  
 Db 269 AlaIleuAsnTrpValSerLeuLeuSerLeuIleAsnAlaGlnIleAsnMetAlaGly 288  
 QY 844 ACCAC 903  
 Db 289 ThrProTyrIleAlaGlnIleGlyThrProLeuLeuGlnIleAspThrAlaLeu 308  
 QY 904 AGCGCCCATCCAC 954  
 Db 309 Thr-----LeuGlnLeuAspAlaGlnGlyLeuLysLeuProIleSerAlaGlnAsn 325  
 QY 955 -----CTCTCTTATCCACACACACACACACACACACACACACACACACAC 1011  
 Db 326 ArgValLeuPheLeuGlyGlyHisAspThrAsnIleAlaAsnIleAlaGlyMetLeuGly 345  
 QY 1012 CTCACTGGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1071  
 Db 346 AlaAspTrpGlnLeuProGlnIleProAspAsnThrProProGlyGlyGlyLeuValPhe 365  
 QY 1072 GAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1131  
 Db 366 GluLeuTrpGlnAsnProAspAsnHisGlnArgTyrValAlaValLysMetPheTyrGln 385  
 QY 1132 ACTTTAC 1188  
 Db 386 ThrMetAspGlnLeuArgAsnAlaGlnLysLeuAspLeuLysAsnAsnProAlaGlyIle 405  
 QY 1189 CTCAACTCCCTGCGACGACATGACACACACACACACACACACACACACAC 1248  
 Db 406 ThrSerValAlaValAlaGlyCysGluAsnAsnGlyAspAspLysLeuGlySerIle 425  
 QY 1249 GCTTACGACCAATCGGAAGACACACACACACACACACACACACACACAC 1296  
 Db 426 ThrPheGlnLysValAlaLysValIleGluSerAlaCysHisIle 441

RESULT 5  
 F90773  
 C:Species: Escherichia coli  
 C:Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #ext\_change 18-Jul-2001  
 C:Accession: F90773  
 F:Hayashi, T.; Marino, K.; Ohnishi, M.; Kurakawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;  
 Kasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
 DNA Res. 8, 11-22, 2001  
 A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genome  
 A:Reference number: A99629; MIM:21156241; PMID:11258766  
 A:Accession: F90773  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-413 <HAY>  
 A:Cross-references: GR:BA000007; PID:BA04561.1; PID:q13360616; GSPLB:GN00154  
 A:Experimental source: strain O157:H7, substrain RMD 050952  
 C:Genetics:  
 A:Gene: ECs1158  
 Alignment Scores:  
 Pred. No.: 2e-48 Length: 413  
 Score: 560.00 Matches: 141  
 Percent Similarity: 50.47% Conservative: 73  
 Best local Similarity: 33.25% Mismatches: 174  
 Query Match: 24.25% Indels: 36  
 DB: 2 Gaps: 12

US-09-866-379a-9 (1-1308) x F9073 (1-413)  
 QY 37 ATTCTGTTAAACCCGCAATCTGCTCATTCGCTGACAGTGAAGTCAG----CTGAAGTTGGAA 94  
 Db 13 ValValLeuLeuAlaSerAsnAlaGlnAlaGlnIleValProIleuLysIleGlnLeuGln 24  
 QY 94 AATNTGTTGATTTGTCAGTGGTATGATGTCGCTGCTCCA----ACCAGGCGACAGAACATG 150  
 Db 33 GlnValLeuMetSerArgHisAsnLeuArgAlaProLeuAlaAsnAsnGlySerVal 52  
 QY 151 ATGACACATCTACCTACACAGCTATGAGGAAAGAGGAGGAGGAGGAGGAGGAGGAG 210  
 Db 53 LeuGlnIleSerThrProAsnLysIleProGluIleAspValProLysIleGlnLeuThr 72  
 QY 211 CCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 270  
 Db 73 ThrLysGlyGlyValLeuGluValTyrMetGlyHisTyrMetArgGluIleLeuAlaGln 92  
 QY 271 GACGGATGCTGCTTAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 330  
 Db 93 GlnGlyMetValLysSerGlyGluCysProProProAspThrValTyrAlaTyrAlaAsn 112  
 QY 331 GTCGACGAGGTATCCGTAAATAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 390  
 Db 113 SerLeuGlnArgThrValAlaIleAlaGlnPheIleThrIleAlaPheProGlyCys 132  
 QY 391 CCAATACCGCTATACCTAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 450  
 Db 133 AspIleProValHisHisIleGlnGlySerGlyThrMetAspProThrPheAsnProVal 152  
 QY 451 AAAATGCTGCTTGTCCCAACTGGATACGCGAGGAGGAGGAGGAGGAGGAGGAGGAG 501  
 Db 153 IleThrAsp-----AspSerAlaAlaPheSerGluIleAlaValAlaAlaMet 168  
 QY 502 GAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 561  
 Db 169 GluLysGluLeuSerLysLeu---GlnLeuThrAspSerTyrGlnLeu----- 183  
 QY 562 CTGACAGGGGCTGCTTAAATTTGCGCAACACAACTTGTGCTTAAAGGAGGAGGAGGAG 621  
 Db 184 LeuGluLysIleValAsnTyrLysAspSerProAlaCys-----LysGluLys 199  
 QY 622 GAAACCTGTCTATTAACGCGACCATTCACATTCACATTCACATTCACATTCACATTC 681  
 Db 200 GlnGlnCysSerLeuValAspCysThrPheSerAlaLysIleGlnIleGlnIle 219  
 QY 682 TCATTAACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 741  
 Db 220 GlyValSerGlyProLeuLysValGlyAsnSerLeuValAspAlaPheThrLeuGlnTyr 239  
 QY 742 CCACAGGCGCAATTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 795  
 Db 240 LysGluLysIlePheProMetAspAlaValAlaIlePheIleLysSerAspThrLeuTyr 259  
 QY 796 AACACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 855  
 Db 260 LysValLeuSerLysLysSerGlyTyrThrAspSerLeuLeuIleSerThrIleVal 279  
 QY 855 GCGGCGACATTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 915  
 Db 280 AlaArgAsnValAlaLysProLeuValSerTyrIleAspLysAlaLeu----- 295  
 QY 915 CCGCAAAACACAGGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 966  
 Db 296 -----ValThrAspArgThrSerAlaProLysIleThrValLeu 408  
 QY 967 GCGGACAGGAGTAAATCTGCGCAAAATCTGCGGCGGCGGCGGCGGCGGCGGCGGCG 1024  
 Db 309 ValGlyHisAspSerAsnIleAlaSerLeuLeuThrAlaLeuAspPheLysProTyrGln 428  
 QY 1024 CTTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1084







Alignment Scores:  
 Pred. No.: 2,41e-37 Length: 413  
 Score: 547.00 Matches: 138  
 Percent Similarity: 49.98% Conservative: 66  
 Best Local Similarity: 33.74% Mismatches: 181  
 Query Match: 23.69% Indels: 24  
 DB: 2 Gaps: 10

US-09-866-379a-9 (1-1308) x AG0632 (1-413)

QY 64 GCTCAGAGTACGCGGAC--CTGAAAGTGGAAAGTGGTCAATCTCAGTCCGTCATCGT 120  
 DB 22 ALaclatThrProGluGlyTyrglnleuglnlnValLeuMetMetSerArgHisasn 41  
 QY 121 GTATGTCGTCA---ATCAAAGTCAAGTCAATGATGATAGATGTCAGCCAGACGATGG 177  
 DB 42 LeuArgAlaProLeuAlaAsnAsnGlyAsnValLeuAlaGlnSerThrProAsnAlatrp 61  
 QY 178 GCAACCTGGTGGGTAAACCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 237  
 DB 62 ProAlatrpAspValProGlyGlyGlnleuThrThrLysGlyGlyValLeuGluValtyr 81  
 QY 248 CTGGAATATACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 297  
 DB 82 MetGlyHisTyrrArqGlnTrpLeuValAlaLeuGlyLeuTleProSerGlyGlyCys 101  
 QY 298 GCGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 357  
 DB 102 ProAlaProAspThrValTyralaTyralaAsnSerLeuGlnArqThrValAlaThrAla 121  
 QY 358 GAAGCTTCG 417  
 DB 122 GluPheBleThrSerAlaPheProGlyCysAspIleProValHisGluClnLys 141  
 QY 418 ACGTCCATCCGATCGTATTTAACTCTCTAAACAGTGGCTTGGCAACTGGATTAAC 477  
 DB 142 MetGlyThrMetAspProThrPheAsnProValIleThrAspAspSerAlaAlaPheArq 161  
 QY 478 GCGAAGCTGACTAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 537  
 DB 162 GluGlnAlaValGlnAla---MotGlyLysAlaArgSerGlnleu----- 175  
 QY 538 CATATCAAGCGGCTTCGGAACCTGGAACGGGTGCTTAATTTCCGCAATCAACATCG 597  
 DB 176 HisLeuAspGlnSerTyrlleStenLeuGlnleuThrHisTyrglnAspSerProSer 195  
 QY 598 TGGCTTAAAGTGAAGAACAG 657  
 DB 196 Cys-----LysGlnLysHisClnCysSerLeuTleAspAlaLysAspThrPhe 211  
 QY 658 CTAAAGTGAAGCGGAGACTGTGTGATTACCGGTGGGTAAAGCTGGCATCAATGCTG 717  
 DB 212 SerAlaAsnTyrglnGlnThrProGlyValGlnIleProLeuLysValGlyAsnSerLeu 231  
 QY 718 AGCAGATATTCCTTCGCAACAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 771  
 DB 242 ValAspAlaPheThrSerTyrlleTyrlleTyrllePheProLeuValAlaValTyrGly 251  
 QY 772 AGCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 831  
 DB 252 GlyIleHisThrAspArgGlnTrpLysValLeuSerTyrlleLysAsnGlyTyrglnAsp 271  
 QY 842 TGTATCAAGCGAGAGTGTGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 891  
 DB 272 SerLeuPheThrSerProThrValAlaAlaAsnValAlaAlaProLeuValLysTyrlle 291  
 QY 892 AAGACAGCTTGAAGCGGATCCAGCGCAAAACAGGCTATGCTGTGATCAATACCACT 951  
 DB 292 AspLysValLeuVal-----AlaGluArgValSerAlaPro--- 304  
 QY 952 TCAGTGGTGTATATGGCGAGACAGATACATAATCTGGCAATCTGGCGCGCACTGGAG 1011  
 DB 93 GlnAsnLysLeuLeuAlaAspLeu-----CysProThrSerAsnIle 1027

DB 304 LysValThrValLeuValGlyHisAspSerAsnIleAlaSerLeuLeuThrAlaLeuAsp 1027  
 QY 1012 CTAAAGCTTCTGAGAGCTTCAGAGTAAAGTAAATAAATAAATAAATAAATAAATAA 1098  
 DB 324 PheLysProTyrglnLeuHisAspGlnTyrlleAlaThrProIleThrGlyGlnLeuVal 1143  
 QY 1069 TTTGAAAGTCTGGTCTGGTAAAGTAAATAAATAAATAAATAAATAAATAAATAAAT 1128  
 DB 344 PheGlnArgTrpHisAspGlyAsnAlaAsnAlaAspLeuMetLysIleLeuTyrlleVal 1188  
 QY 1129 CAGATTTTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 1188  
 DB 364 GlnSerAlaArgGlnLeuArgAsnAlaGlnAlaLeuThrLeuLysSerProAlaAlaArg 1188  
 QY 1189 GTGAAATGACCCCTGGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 1248  
 DB 384 ValThrLeuGluLeuLysGlyCysProVal-AspAlaAsnLysLysProLeuAsp 1248  
 QY 1249 GGTTTTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 1275  
 DB 404 LysPheAspAsnValMetAsnThrAla 1411

RESULT 9  
 S25627  
 glucose-1-phosphatase (EC 3.1.3.10) precursor - Providencia rettgeri  
 C.Studies: Providencia rettgeri  
 G.Date: 25-Feb-1994 #sequence\_revision 10-Nov-1995 #text\_change 06-Oct-1999  
 G.Accession: S25627  
 R.Riccio, M.L. Chiesurin, A. Lombardi, G. Satta, G.  
 submitted to the EMBL Data Library, September 1992  
 A:Reference number: S25627  
 A:Accession: S25627  
 A:Molecule type: DNA  
 A:Residues: 1,417<R3C>  
 A:Cross-references: EMBL:X88201; NID:045771; PIR:CAA4288.1; PIR:045772  
 A:Experimental source: strain pw7  
 C:Genetics:  
 A:Gene: asp  
 C:Keywords: periplasmic space; phosphohistidine; phosphoprotein; phosphor  
 E:1-25/Domain; signal sequence #status predicted -sig  
 F:30-417/Product; glucose-1-phosphatase #status predicted -MAT  
 F:42/Active site; His (phosphohistidine intermediate) #status predicted  
 F:314/Active site; His #status predicted

Alignment Scores:  
 Pred. No.: 1,36e-36 Length: 417  
 Score: 548.00 Matches: 140  
 Percent Similarity: 50.24% Conservative: 84  
 Best Local Similarity: 30.17% Mismatches: 175  
 Query Match: 23.40% Indels: 0  
 DB: 2 Gaps: 11

US-09-866-379a-9 (1-1308) x S25627 (1-417)

QY 28 TCTCTTCGATTCGCTTAACTGAGTAAATGATGATGATGATGATGATGATGATGATGAT 107  
 DB 14 AlaLeuPheAlaProIleAlaProThrMetAlaSerThrAspAsnGlnAlaAspMetVal 107  
 QY 88 CTGCAAACTGGGTCATTCGATGATGATGATGATGATGATGATGATGATGATGATGAT 147  
 DB 33 LeuAspGlnValLeuValLeuSerArgHisAsnLeuAlaThrProLeuValAsnThrleu 147  
 QY 148 CTGATGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 207  
 DB 54 TleuThrGlnValThrAspLysTrpProAspTrpAspLysSerTyrlleProGly 207  
 QY 208 ACAGCGGCGTGGTGGTAAATGATGATGATGATGATGATGATGATGATGATGATGATG 268  
 DB 73 ThrThrGlnGlyAlaLeuLeuValTyrlleMetGlyHisTyrllePheAlaLeuThrPheAsp 268  
 QY 259 -----GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 312  
 DB 93 GlnAsnLysLeuLeuAlaAspLeu-----CysProThrSerAsnIle 312



QY 709 TCAATGCTGACGGACATATTTCTGCTGCAACAACACACAGCAATGCGGAGCCGGGGTGG 768  
 Db 224 SerGlyValThrGluSerLeuMetAlaTrpAlaAspGlyArqAspPheAlaGlyLeu 243  
 QY 769 GGA---AGATATCATGCTATTCACACCTACCTGCAACACCTGCTGCAAGTTTGTCAATGCGGCA 825  
 Db 244 GlyTrpLysSerLeuAspGluGluAlaLeuThrAlaSerPheLeuHisGluAlaGlu 263  
 QY 826 TTGATTGTGATGAAGCGGACAGAGTTGGCGGCGACCGCGGACCGCGTATTATGAT 885  
 Db 264 PheAspLeuArqLeuAlaTrpProTyrValAlaArqThrLeuAlaGlyHisLeuAlaAsp 283  
 QY 886 TGCATTAAGACAGAGCTTGAGAGCGCATCCACAGCGCAAAAACAGGCGATGGTGCACATTA 945  
 Db 284 ArgLeuAlaAlaThrLeuArq---AspGlyAlaAlaAlaIleGly---ProVal 299  
 QY 946 GTCATTCAGTGTCTTATGCGGACACACATATGCGCAAAATGCGGCGGCGCA 1005  
 Db 300 AspAlaAlaLeuValIleLeuAlaGlyHisAspGlyThrLeuAlaSerLeuGlyGlyLeu 319  
 QY 1006 CTGAGATTAACATGAGAGCTTCGAGT---CATGCGCATATATATGCGGCGGCA 1062  
 Db 320 LeuArqMetGluTrpThrLeuProGlyTyrGlnProAsnGlnIleGlnProGlyGlyAla 339  
 QY 1063 CTGCTGTTGAATGCTGCGGCGGCTAAAGGATAACACCGAGTGGATTCAGGTTTCGCTG 1122  
 Db 340 LeuValPheGluArqTrpArqArq---AspAspGlyValArqValArqValArqPhe 358  
 QY 1123 GTCATTCAGACTTACAGCAGATGCTGATGAATAACGCGCGTGCATTAAATACCGCGCC 1182  
 Db 359 ThrGlyGlnSerLeuSerGlnLeuArqAsnMetThrAlaLeuAspAlaTyrThrProPro 378  
 QY 1183 GAGAGATTAAGATGAGCTGCTGAGGATGCAAGAGCAAAATGCGAGGCGATGTGTTCG 1242  
 Db 379 LeuSerAlaProValPheValGlnGlyCysGlyThrAlaThrProAlaPheAspCysArq 398  
 QY 1243 TGGACAGCTTTTACGAAATGAGCAATGCAAGCA 1275  
 Db 399 LeuGluAspPheGluThrValValArqGlyAla 409  
 RESULT 11  
 116058  
 hypothetical protein F14011.1 Caenorhabditis elegans  
 c:Species: caenorhabditis elegans  
 c:date: 20 Sep 1999 #sequence\_revision 20-Sep-1999 #text\_change 23-Mar-2001  
 c:Accession: T16058  
 R:Blatt, L.  
 submitted to the EMBL data library, November 1995  
 A:Description: the sequence of C. elegans cosmid F14011.  
 A:Reference number: S69020  
 A:Accession: T16058  
 A:Gene: F14011.1  
 A:Molecule type: DNA  
 A:References: 1-416 EMBL  
 A:Cross references: EMBL:040939; NID:q1072175; PID:q1072178; P1DN:AAA81702.1; CESP:F1304  
 c:Genes:  
 A:Exons: 234, 594, 984, 1043, 1334, 1443, 1703, 1813, 1913, 2013, 2113, 2213, 2313, 2413, 2513, 2613, 2713, 2813, 2913, 3013, 3113, 3213, 3313, 3413, 3513, 3613, 3713, 3813, 3913, 4013, 4113, 4213, 4313, 4413, 4513, 4613, 4713, 4813, 4913, 5013, 5113, 5213, 5313, 5413, 5513, 5613, 5713, 5813, 5913, 6013, 6113, 6213, 6313, 6413, 6513, 6613, 6713, 6813, 6913, 7013, 7113, 7213, 7313, 7413, 7513, 7613, 7713, 7813, 7913, 8013, 8113, 8213, 8313, 8413, 8513, 8613, 8713, 8813, 8913, 9013, 9113, 9213, 9313, 9413, 9513, 9613, 9713, 9813, 9913, 10013, 10113, 10213, 10313, 10413, 10513, 10613, 10713, 10813, 10913, 11013, 11113, 11213, 11313, 11413, 11513, 11613, 11713, 11813, 11913, 12013, 12113, 12213, 12313, 12413, 12513, 12613, 12713, 12813, 12913, 13013, 13113, 13213, 13313, 13413, 13513, 13613, 13713, 13813, 13913, 14013, 14113, 14213, 14313, 14413, 14513, 14613, 14713, 14813, 14913, 15013, 15113, 15213, 15313, 15413, 15513, 15613, 15713, 15813, 15913, 16013, 16113, 16213, 16313, 16413, 16513, 16613, 16713, 16813, 16913, 17013, 17113, 17213, 17313, 17413, 17513, 17613, 17713, 17813, 17913, 18013, 18113, 18213, 18313, 18413, 18513, 18613, 18713, 18813, 18913, 19013, 19113, 19213, 19313, 19413, 19513, 19613, 19713, 19813, 19913, 20013, 20113, 20213, 20313, 20413, 20513, 20613, 20713, 20813, 20913, 21013, 21113, 21213, 21313, 21413, 21513, 21613, 21713, 21813, 21913, 22013, 22113, 22213, 22313, 22413, 22513, 22613, 22713, 22813, 22913, 23013, 23113, 23213, 23313, 23413, 23513, 23613, 23713, 23813, 23913, 24013, 24113, 24213, 24313, 24413, 24513, 24613, 24713, 24813, 24913, 25013, 25113, 25213, 25313, 25413, 25513, 25613, 25713, 25813, 25913, 26013, 26113, 26213, 26313, 26413, 26513, 26613, 26713, 26813, 26913, 27013, 27113, 27213, 27313, 27413, 27513, 27613, 27713, 27813, 27913, 28013, 28113, 28213, 28313, 28413, 28513, 28613, 28713, 28813, 28913, 29013, 29113, 29213, 29313, 29413, 29513, 29613, 29713, 29813, 29913, 30013, 30113, 30213, 30313, 30413, 30513, 30613, 30713, 30813, 30913, 31013, 31113, 31213, 31313, 31413, 31513, 31613, 31713, 31813, 31913, 32013, 32113, 32213, 32313, 32413, 32513, 32613, 32713, 32813, 32913, 33013, 33113, 33213, 33313, 33413, 33513, 33613, 33713, 33813, 33913, 34013, 34113, 34213, 34313, 34413, 34513, 34613, 34713, 34813, 34913, 35013, 35113, 35213, 35313, 35413, 35513, 35613, 35713, 35813, 35913, 36013, 36113, 36213, 36313, 36413, 36513, 36613, 36713, 36813, 36913, 37013, 37113, 37213, 37313, 37413, 37513, 37613, 37713, 37813, 37913, 38013, 38113, 38213, 38313, 38413, 38513, 38613, 38713, 38813, 38913, 39013, 39113, 39213, 39313, 39413, 39513, 39613, 39713, 39813, 39913, 40013, 40113, 40213, 40313, 40413, 40513, 40613, 40713, 40813, 40913, 41013, 41113, 41213, 41313, 41413, 41513, 41613, 41713, 41813, 41913, 42013, 42113, 42213, 42313, 42413, 42513, 42613, 42713, 42813, 42913, 43013, 43113, 43213, 43313, 43413, 43513, 43613, 43713, 43813, 43913, 44013, 44113, 44213, 44313, 44413, 44513, 44613, 44713, 44813, 44913, 45013, 45113, 45213, 45313, 45413, 45513, 45613, 45713, 45813, 45913, 46013, 46113, 46213, 46313, 46413, 46513, 46613, 46713, 46813, 46913, 47013, 47113, 47213, 47313, 47413, 47513, 47613, 47713, 47813, 47913, 48013, 48113, 48213, 48313, 48413, 48513, 48613, 48713, 48813, 48913, 49013, 49113, 49213, 49313, 49413, 49513, 49613, 49713, 49813, 49913, 50013, 50113, 50213, 50313, 50413, 50513, 50613, 50713, 50813, 50913, 51013, 51113, 51213, 51313, 51413, 51513, 51613, 51713, 51813, 51913, 52013, 52113, 52213, 52313, 52413, 52513, 52613, 52713, 52813, 52913, 53013, 53113, 53213, 53313, 53413, 53513, 53613, 53713, 53813, 53913, 54013, 54113, 54213, 54313, 54413, 54513, 54613, 54713, 54813, 54913, 55013, 55113, 55213, 55313, 55413, 55513, 55613, 55713, 55813, 55913, 56013, 56113, 56213, 56313, 56413, 56513, 56613, 56713, 56813, 56913, 57013, 57113, 57213, 57313, 57413, 57513, 57613, 57713, 57813, 57913, 58013, 58113, 58213, 58313, 58413, 58513, 58613, 58713, 58813, 58913, 59013, 59113, 59213, 59313, 59413, 59513, 59613, 59713, 59813, 59913, 60013, 60113, 60213, 60313, 60413, 60513, 60613, 60713, 60813, 60913, 61013, 61113, 61213, 61313, 61413, 61513, 61613, 61713, 61813, 61913, 62013, 62113, 62213, 62313, 62413, 62513, 62613, 62713, 62813, 62913, 63013, 63113, 63213, 63313, 63413, 63513, 63613, 63713, 63813, 63913, 64013, 64113, 64213, 64313, 64413, 64513, 64613, 64713, 64813, 64913, 65013, 65113, 65213, 65313, 65413, 65513, 65613, 65713, 65813, 65913, 66013, 66113, 66213, 66313, 66413, 66513, 66613, 66713, 66813, 66913, 67013, 67113, 67213, 67313, 67413, 67513, 67613, 67713, 67813, 67913, 68013, 68113, 68213, 68313, 68413, 68513, 68613, 68713, 68813, 68913, 69013, 69113, 69213, 69313, 69413, 69513, 69613, 69713, 69813, 69913, 70013, 70113, 70213, 70313, 70413, 70513, 70613, 70713, 70813, 70913, 71013, 71113, 71213, 71313, 71413, 71513, 71613, 71713, 71813, 71913, 72013, 72113, 72213, 72313, 72413, 72513, 72613, 72713, 72813, 72913, 73013, 73113, 73213, 73313, 73413, 73513, 73613, 73713, 73813, 73913, 74013, 74113, 74213, 74313, 74413, 74513, 74613, 74713, 74813, 74913, 75013, 75113, 75213, 75313, 75413, 75513, 75613, 75713, 75813, 75913, 76013, 76113, 76213, 76313, 76413, 76513, 76613, 76713, 76813, 76913, 77013, 77113, 77213, 77313, 77413, 77513, 77613, 77713, 77813, 77913, 78013, 78113, 78213, 78313, 78413, 78513, 78613, 78713, 78813, 78913, 79013, 79113, 79213, 79313, 79413, 79513, 79613, 79713, 79813, 79913, 80013, 80113, 80213, 80313, 80413, 80513, 80613, 80713, 80813, 80913, 81013, 81113, 81213, 81313, 81413, 81513, 81613, 81713, 81813, 81913, 82013, 82113, 82213, 82313, 82413, 82513, 82613, 82713, 82813, 82913, 83013, 83113, 83213, 83313, 83413, 83513, 83613, 83713, 83813, 83913, 84013, 84113, 84213, 84313, 84413, 84513, 84613, 84713, 84813, 84913, 85013, 85113, 85213, 85313, 85413, 85513, 85613, 85713, 85813, 85913, 86013, 86113, 86213, 86313, 86413, 86513, 86613, 86713, 86813, 86913, 87013, 87113, 87213, 87313, 87413, 87513, 87613, 87713, 87813, 87913, 88013, 88113, 88213, 88313, 88413, 88513, 88613, 88713, 88813, 88913, 89013, 89113, 89213, 89313, 89413, 89513, 89613, 89713, 89813, 89913, 90013, 90113, 90213, 90313, 90413, 90513, 90613, 90713, 90813, 90913, 91013, 91113, 91213, 91313, 91413, 91513, 91613, 91713, 91813, 91913, 92013, 92113, 92213, 92313, 92413, 92513, 92613, 92713, 92813, 92913, 93013, 93113, 93213, 93313, 93413, 93513, 93613, 93713, 93813, 93913, 94013, 94113, 94213, 94313, 94413, 94513, 94613, 94713, 94813, 94913, 95013, 95113, 95213, 95313, 95413, 95513, 95613, 95713, 95813, 95913, 96013, 96113, 96213, 96313, 96413, 96513, 96613, 96713, 96813, 96913, 97013, 97113, 97213, 97313, 97413, 97513, 97613, 97713, 97813, 97913, 98013, 98113, 98213, 98313, 98413, 98513, 98613, 98713, 98813, 98913, 99013, 99113, 99213, 99313, 99413, 99513, 99613, 99713, 99813, 99913, 100013, 100113, 100213, 100313, 100413, 100513, 100613, 100713, 100813, 100913, 101013, 101113, 101213, 101313, 101413, 101513, 101613, 101713, 101813, 101913, 102013, 102113, 102213, 102313, 102413, 102513, 102613, 102713, 102813, 102913, 103013, 103113, 103213, 103313, 103413, 103513, 103613, 103713, 103813, 103913, 104013, 104113, 104213, 104313, 104413, 104513, 104613, 104713, 104813, 104913, 105013, 105113, 105213, 105313, 105413, 105513, 105613, 105713, 105813, 105913, 106013, 106113, 106213, 106313, 106413, 106513, 106613, 106713, 106813, 106913, 107013, 107113, 107213, 107313, 107413, 107513, 107613, 107713, 107813, 107913, 108013, 108113, 108213, 108313, 108413, 108513, 108613, 108713, 108813, 108913, 109013, 109113, 109213, 109313, 109413, 109513, 109613, 109713, 109813, 109913, 110013, 110113, 110213, 110313, 110413, 110513, 110613, 110713, 110813, 110913, 111013, 111113, 111213, 111313, 111413, 111513, 111613, 111713, 111813, 111913, 112013, 112113, 112213, 112313, 112413, 112513, 112613, 112713, 112813, 112913, 113013, 113113, 113213, 113313, 113413, 113513, 113613, 113713, 113813, 113913, 114013, 114113, 114213, 114313, 114413, 114513, 114613, 114713, 114813, 114913, 115013, 115113, 115213, 115313, 115413, 115513, 115613, 115713, 115813, 115913, 116013, 116113, 116213, 116313, 116413, 116513, 116613, 116713, 116813, 116913, 117013, 117113, 117213, 117313, 117413, 117513, 117613, 117713, 117813, 117913, 118013, 118113, 118213, 118313, 118413, 118513, 118613, 118713, 118813, 118913, 119013, 119113, 119213, 119313, 119413, 119513, 119613, 119713, 119813, 119913, 120013, 120113, 120213, 120313, 120413, 120513, 120613, 120713, 120813, 120913, 121013, 121113, 121213, 121313, 121413, 121513, 121613, 121713, 121813, 121913, 122013, 122113, 122213, 122313, 122413, 122513, 122613, 122713, 122813, 122913, 123013, 123113, 123213, 123313, 123413, 123513, 123613, 123713, 123813, 123913, 124013, 124113, 124213, 124313, 124413, 124513, 124613, 124713, 124813, 124913, 125013, 125113, 125213, 125313, 125413, 125513, 125613, 125713, 125813, 125913, 126013, 126113, 126213, 126313, 126413, 126513, 126613, 126713, 126813, 126913, 127013, 127113, 127213, 127313, 127413, 127513, 127613, 127713, 127813, 127913, 128013, 128113, 128213, 128313, 128413, 128513, 128613, 128713, 128813, 128913, 129013, 129113, 129213, 129313, 129413, 129513, 129613, 129713, 129813, 129913, 130013, 130113, 130213, 130313, 130413, 130513, 130613, 130713, 130813, 130913, 131013, 131113, 131213, 131313, 131413, 131513, 131613, 131713, 131813, 131913, 132013, 132113, 132213, 132313, 132413, 132513, 132613, 132713, 132813, 132913, 133013, 133113, 133213, 133313, 133413, 133513, 133613, 133713, 133813, 133913, 134013, 134113, 134213, 134313, 134413, 134513, 134613, 134713, 134813, 134913, 135013, 135113, 135213, 135313, 135413, 135513, 135613, 135713, 135813, 135913, 136013, 136113, 136213, 136313, 136413, 136513, 136613, 136713, 136813, 136913, 137013, 137113, 137213, 137313, 137413, 137513, 137613, 137713, 137813, 137913, 138013, 138113, 138213, 138313, 138413, 138513, 138613, 138713, 138813, 138913, 139013, 139113, 139213, 139313, 139413, 139513, 139613, 139713, 139813, 139913, 140013, 140113, 140213, 140313, 140413, 140513, 140613, 140713, 140813, 140913, 141013, 141113, 141213, 141313, 141413, 141513, 141613, 141713, 141813, 141913, 142013, 142113, 142213, 142313, 142413, 142513, 142613, 142713, 142813, 142913, 143013, 143113, 143213, 143313, 143413, 143513, 143613, 143713, 143813, 143913, 144013, 144113, 144213, 144313, 144413, 144513, 144613, 144713, 144813, 144913, 145013, 145113, 145213, 145313, 145413, 145513, 145613, 145713, 145813, 145913, 146013, 146113, 146213, 146313, 146413, 146513, 146613, 146713, 146813, 146913, 147013, 147113, 147213, 147313, 147413, 147513, 147613, 147713, 147813, 147913, 148013, 148113, 148213, 148313, 148413, 148513, 148613, 148713, 148813, 148913, 149013, 149113, 149213, 149313, 149413, 149513, 149613, 149713, 149813, 149913, 150013, 150113, 150213, 150313, 150413, 150513, 150613, 150713, 150813, 150913, 151013, 151113, 151213, 151313, 151413, 151513, 151613, 151713, 151813, 151913, 152013, 152113





















```

Db 201 LeuValSerAlaAspValSerLeuThrGlnAlaLeuProSerGlnLeu 220
QY 661 AAGTCAGGCGGAGCTGTGTCTCATTAAGCGGTGGGTGAAGCCTGGCATCAATGCTGACG 720
Db 221 LysValSerAlaAspAsnValSerLeuThrGlyAlaValSerLeuAlaSerMetLeuThr 240
QY 721 GAGATATTCTCTCGCAACAGCAGCAGGGAATGCGGAGCGGCGGCGGCGGCGGCGGCGG 780
Db 241 GlnThrPheLeuLeuGlnGlnAlaGlnGlyMetProGlyTrpGlyValGlnThr 260
QY 781 GATTCACAGCAGCGGAGACCTGCTCAATTTTCATTAACGCGCAATTTGCTGACAA 840
Db 261 AspSerHisGlnTrpAsnThrLeuLeuSerLeuHisAsnAlaGlnThrLeuLeuGln 280
QY 841 CGACATGCAACAGCTGTGCGGAGCGGCGGTAAGGCTGATTAATTAATTAATTAATTAATTA 900
Db 281 ArgThrProGluValAlaArgSerArgAlaThrProLeuLeuAspLeuLeuLysThrAla 360
QY 901 TTGACGCGCATTCACACGCAAAAACAGCGCTATGCTGTCACATTAGCCACTTTCAGTCTG 960
Db 301 LeuThrProHisProProGlnGlyAlaGlyValThrLeuProThrSerValLeu 320
QY 961 TTATGCGCGGACACATAGCTAATCTGTAATCTGTAATCTGTAATCTGTAATCTGTAATCTG 1020
Db 321 PheThrAlaGlyHisAspThrAsnLeuAlaAsnLeuGlyGlyAlaLeuGlnLeuAsnTrp 340
QY 1021 ACCTTCGCGGTCAGCGGATAAACAGCGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1080
Db 341 ThrLeuProGlyGlnProAspAsnThrProProGlyGlyGlnLeuValPheGlnArgTrp 360
QY 1081 CCGCGCGCTAAGCGATACACATTCATTCAGCTTTGCGTGGCTGCCAGACTTTACAG 1140
Db 361 ArgArgLeuSerAspAsnSerGlnTrpLeuValSerLeuValPheGlnThrLeuLeu 380
QY 1141 CACATGCGCTAATAACAGCGGCTTTTATTAAATAGCGCGCGCGGAGAGGAGAACTGAC 1200
Db 381 GluMetArgAspLysThrProLeuSerLeuAsnThrProProGlyGlnValLysLeuThr 400
QY 1201 CTGCTAGATGTCAAGAGCGAAATGCGCAGCGCATGTGCTGCTGGCGAGCTTTTACGCAA 1260
Db 401 LeuAlaGlyCysGlnGluArgAsnAlaGlnGlyMetCysSerLeuAlaGlyPheThrGln 420
QY 1261 ATGCTGTAATGACACAGCATACCGCGGTGGTGGAGTTG 1296
Db 421 ThrValAsnGlnAlaAlaGlnThrAlaCysSerLeu 432

```

RESULT 6  
QRRKDS  
ID QRRKDS FREE-EMANRY: 487: 432 AA.  
AC QRRKDS:  
DT 01-JUN-2002 (TEMBLrel, 21, created)  
DI 01-JUN-2002 (TEMBLrel, 21, last sequence update)  
DI 01-JUN-2002 (TEMBLrel, 21, last annotation update)  
DE Porc plasmic phosphoanhydride phosphohydrolase.  
GN ALPA.  
OS Escherichia coli.  
OC Bacteria, Proteobacteria, gamma subdivision, Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID:562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE 90368616; PubMed 2168485;  
RA Bussa J., Marek C., Poppe P., Leli.  
RT "The complete nucleotide sequence of the Escherichia coli gene appA  
reveals significant homology between pH 2.5 acid phosphatase and  
glucose-1-phosphatase."  
RL J. Bacteriol. 172:5497-5500(1990).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE 94054596; PubMed 1429641;  
RA Ostman K., Harms E.H., Stevis F.E., Kuebel R., Zhou M.M.,  
Van Etten R.L.  
RT "Overexpression, site directed mutagenesis, and mechanism of

```

RT Escherichia coli acid phosphatase."
RL J. Biol. Chem. 267:22836-22846(1992).
DR EMBL: 104375; AAA00007.1; .
KW Hydrolase.
SQ SEQUENCE 432 AA; 46990 MW; 95DF994EA9A1A47C CRC64;

```

Alignment Scores:  
Prod. No.: 1-170-177 Length: 432  
Score: 2172.00 Matches: 424  
Percent Similarity: 97.92% Conservation: 0  
Best Local Similarity: 97.92% Mismatches: 0  
Query Match: 94.07% Indels: 0  
Gaps: 0

US-09-866-379A-9 (1-1308) X QRRKDS (1-432)

```

QY 1 ATGAAAAGATATTAATTCGCAATTTTATCTCTCTGTAATTCGTAATTCGTAATTCGTAATTCG 60
Db 1 MetLysAlaIleLeuIleProPheLeuSerLeuLeuIleProIleProIleProIleProIleProIle 20
QY 61 TTGATTCACATGACGCGGAGCTCAAGCTGGAAGATCTGGTGTGATGTCATGCTGATGTCATGCT 120
Db 21 PheAlaGlnSerGlnProGlnLeuLysLeuGlnSerValValIleValSerArgHisGly 40
QY 121 GTGGTCTCTCAACCAAGGTCACGCAATTCATGTCATGTCATGTCATGTCATGTCATGTCATGTC 180
Db 41 ValArgAlaProThrLysAlaThrGlnLeuMetGlnAspValThrProAspAlaThrPro 60
QY 181 ACTGGCGGTTAAACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240
Db 61 ThrProValLysLeuGlyTrpLeuThrProArgLysLeuIleIleIleIleIleIleIleIleIleIle 80
QY 241 GAGCATCTGCGGCTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 300
Db 81 GlyHisLysTrpGlnArgGlnArgLeuValAlaAspGlyLeuLeuAlaLysGlySerGlySerPro 100
QY 301 CAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360
Db 101 GluSerGlyGlnValAlaIleIleValAspValAspLeuArgThrArgLysThrGlyThr 120
QY 361 GCTCTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
Db 121 AlaThrAlaAlaGlyLeuAlaProAspCysAlaIleThrValHisThrThrAlaAspThr 140
QY 421 TCAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 480
Db 141 SerSerProAspProLeuPheAspTrpLeuLysThrGlyValGlyGlnLeuAspAsnAla 160
QY 481 AACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
Db 161 AsnValThrAspAlaIleLeuSerArgAlaGlyLysSerIleAlaAspPheThrGlyHis 180
QY 541 TATCAAAAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
Db 181 ArgLeuThrAlaPheArgGlnLeuGlnArgValLeuAsnThrProThrSerAsnLeuGly 200
QY 601 CTCAAAAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
Db 201 LeuLysArgLeuLysGlnAspGlnArgCysSerLeuThrGlnAlaLeuProSerThrLeu 220
QY 661 AAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720
Db 221 LysValSerAlaAspAsnValSerLeuThrGlyAlaValSerLeuAlaSerMetLeuThr 240
QY 721 GAGATATTCTCTCGCAACAGCAGCAGGGAATGCGGAGCGGCGGCGGCGGCGGCGGCGGCGG 780
Db 241 GlnThrPheLeuLeuGlnGlnAlaGlnGlyMetProGlyTrpGlyValGlnThr 260
QY 781 GATTCACAGCAGCGGAGACCTGCTCAATTTTCATTAACGCGCAATTTGCTGACAA 840
Db 261 AspSerHisGlnTrpAsnThrLeuLeuSerLeuHisAsnAlaGlnThrLeuLeuGln 280
QY 841 CGACATGCAACAGCTGTGCGGAGCGGCGGTAAGGCTGATTAATTAATTAATTAATTAATTAATTA 900
Db 281 ArgThrProGluValAlaArgSerArgAlaThrProLeuLeuAspLeuLeuLysThrAla 360
QY 901 TTGACGCGCATTCACACGCAAAAACAGCGCTATGCTGTCACATTAGCCACTTTCAGTCTG 960
Db 301 LeuThrProHisProProGlnGlyAlaGlyValThrLeuProThrSerValLeu 320
QY 961 TTATGCGCGGACACATAGCTAATCTGTAATCTGTAATCTGTAATCTGTAATCTGTAATCTG 1020
Db 321 PheThrAlaGlyHisAspThrAsnLeuAlaAsnLeuGlyGlyAlaLeuGlnLeuAsnTrp 340
QY 1021 ACCTTCGCGGTCAGCGGATAAACAGCGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1080
Db 341 ThrLeuProGlyGlnProAspAsnThrProProGlyGlyGlnLeuValPheGlnArgTrp 360
QY 1081 CCGCGCGCTAAGCGATACACATTCATTCAGCTTTGCGTGGCTGCCAGACTTTACAG 1140
Db 361 ArgArgLeuSerAspAsnSerGlnTrpLeuValSerLeuValPheGlnThrLeuLeu 380
QY 1141 CACATGCGCTAATAACAGCGGCTTTTATTAAATAGCGCGCGGAGAGGAGAACTGAC 1200
Db 381 GluMetArgAspLysThrProLeuSerLeuAsnThrProProGlyGlnValLysLeuThr 400
QY 1201 CTGCTAGATGTCAAGAGCGAAATGCGCAGCGCATGTGCTGCTGGCGAGCTTTTACGCAA 1260
Db 401 LeuAlaGlyCysGlnGluArgAsnAlaGlnGlyMetCysSerLeuAlaGlyPheThrGln 420

```

```

Db 281 ArqThrProGluValAlaArqSerArqAlaThrProLeuLeuAspLeuLeuLysThrAla 300
QY 901 TTGACGCGCCACACCAACGCAAAACAGCGGTATGGTGTGCACATTACCCACTTCAGTGTG 960
Db 301 LeuThrProHisProProGluGlySerAlaGlyGlyValThrLeuProThrSerValLeu 320
QY 902 TTTATCGGCGACACGATATCAATTAATGAGGAAAGGAGGAGGAGGAGGAGGAGGAGGAG 1020
Db 321 TheHeLaGlyAlaAspThrAspLeuAlaAspLeuAlaAspLeuAlaAspLeuAla 340
QY 1021 AGGCTTCGCGGTCAGCGGAGAAATACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1080
Db 341 ThrLeuProGlyGluProAspAspSerProGluGlyGlyGluValPheGluArgTrp 360
QY 1081 GGTGCGGCAAGGAGATACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1140
Db 361 ArqArqLeuSerAspAspSerGlnTrpIleGlnValSerLeuValPheGlnThrLeuGln 380
QY 1141 GATATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1200
Db 381 GlnMetArgAspLysThrProLeuSerLeuAsnThrProGluGlyGlyValLysLeuThr 400
QY 1201 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1260
Db 401 LeuAlaGlyCysGluGlnArgAspAlaGlnGlyMetCysSerLeuAlaGlyPheThrGln 420
QY 1261 AFGCTCAATGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1296
Db 421 lleValAsnGluAlaArgIleProAlaCysSerLeu 432
RESULT 7
Q8XC29
ID Q8XC29 PRELIMINARY: PRT: 434 AA.
AC Q8XC29;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Phosphoenolpyruvate phosphorylase, pH 2.5 acid phosphatase,
DE periplasmic.
GN APPA OR 21397 OR E3811d6.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=83334;
RX SEQUENCE FROM N. A.
RC MEDLINE=0157:H7 / EDI933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Pose N.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Postai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grotbeck K.J., Davis N.W., Lim A., Dimalanta E.T., Potamoudis K.,
RA Apolara J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch P.A., Rattiner P.P.;
RT "Genome Sequence of Enterobacteriaceae Escherichia coli O157:H7";
RL Nature 406 529-534 (2001);
RP SEQUENCE FROM N. A.
RC STRAIN=0157:H7 / RMD 0509952;
RX MEDLINE 21152331; PubMed 12258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Yurata T., Tanaka M., Tobe T.,
RA Tada T., Takani H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kubara S., Shiba T., Hattori M., Shinajawa H.,
RT "Complete genome sequence of enterohemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K12";
RL DNA Res. 8:11-22 (2001);
DR EMBL: AF005232; AAC5528.1; ALU:IN17;
DR EMBL: AF002554; BAB34559.1;
DR InterPro: IPR000560; HisAc-phosphatase.
DR Pfam: PF00328; acid phosphatase 1.
DR PROSITE: P500616; HIS_ACID_PHOSPHAT_1: 1.

```

```

DR PROSITE: P500778; HIS_ACID_PHOSPHAT_2: 1.
KW Complete proteome.
SQ SEQUENCE 434 AA: 47337 MW: F197DFD1869F9C4 CR:64;
Alignment Scores:
Pred. No.: 2 95%-175 Length: 434
Score: 2144.60 Matches: 419
Percent Similarity: 96.42 Gaps: 0
Best Local Similarity: 96.54 Mismatches: 13
Query Match: 92.85 Indels: 2
Gaps: 1
US-09-866-379a-9 (1-1308) x Q8XC29 (1-434)
QY 1 ATCAAGGATCTTAATCCATTTTATGCTCTTGATTCGGTTAAGCGCGCAATTCGA 60
Db 1 MetLysAlaIleLeuIleProPheLeuSerLeuLeuIleProLeuThrProGlnSerAla 20
QY 61 TTCGCTCAGAGT-----GAGCGGAGCTCAAGCTGCAAGTGTGTGATTCATTCGT 114
Db 21 PheAlaGlnSerGluProGluProGluLeuLysLeuGluSerValIleValSerArg 40
QY 115 CATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 174
Db 41 HisGlyValArgAlaProThrLysAlaThrGlnLeuMetGluAspValThrProAspAla 60
QY 175 TCGCAACCTGGCGGTAAACTGGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 234
Db 61 TrpLeuSerProValLysGlyTrpLeuThrPheArgGlyGlyGlnLeuLeuAla 80
QY 235 TATCGGACATTACTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 294
Db 81 TyrLeuGlyHisGlyGlnArgGlnArgLeuValAlaAspGlyLeuThrLysLysGly 100
QY 295 TCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 354
Db 101 CysProGlnProGlyGlnValAlaIleLeuAlaAspValAspValArgThrArgLysThr 120
QY 355 GCGCAACCTTCGCGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 414
Db 121 GlyGlnAlaPheAlaGlyGlnAlaPheAspGlyAlaThrValHisThrGlnAla 140
QY 415 GATACGCTGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 474
Db 141 AspThrSerProAspProLeuPheAsnProLeuLysThrGlyValCysGlnLeuAsp 160
QY 475 AATCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 534
Db 161 AsnAlaAsnValThrAspAlaIleLeuSerArgAlaGlyGlySerIleAlaAspPheThr 180
QY 535 GGCATTATCAACGCGGTTCGCAACTGGAACGGGTGCTTAATTTCCGCAATCAAC 594
Db 181 GlyHisArgGlnThrAlaPheAlaGlyGlnLeuGluArgValLeuAspPheProGlnSer 200
QY 595 TTCTGCTTAAAGCTCAGAAATACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 654
Db 201 LeuCysLeuAsnArgCysLysGlnAspGluSerCysSerLeuThrGlnAlaLeuProSer 220
QY 655 CAATCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 714
Db 221 GluLeuLysValSerAlaAspAspValSerLeuThrGlyAlaValIleLeuAlaSerMet 240
QY 715 GTGACGACATATTTCTCTGCAACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 774
Db 241 LeuThrGlnThrPheLeuLeuGlnAlaThrGlyMetProGluProGlyTrpCysArg 260
QY 775 ATCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 834
Db 261 LeuThrAspSerHisGlnThrPheSerLeuLeuSerLeuHisAspAlaGlyPheCysArg 280
QY 835 ATCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 894
Db 281 LeuGlnArgThrProGluValAlaArgSerAlaAlaThrProLeuLeuAspLeuIleMet 900

```







```

QY 1144 ATGAGGATCAAAACGGCGGATGIGATATTAATAACAGCGCGGAGAGAGGTGAACACCCCG 1203
    |||||
DB 369 LeuArASuAlaAspAlaLeuThrLeuGlnAlaProAlaGlnArqValThrLeuGluLeu 488
    |||||
QY 1204 GAGAGAGGCAAGAGGCAAGAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1263
    |||||
DB 389 SerGlyCys---ProLeuAspAlaAspGlyPheCysProMetAspLysPheAspSerVal 407
    |||||
QY 1264 GGAATGAACACA 1275
    |||||
DB 408 LeuAsuGlnAla 411
    |||||

RESULT 10
Q827P1 PRELIMINARY: PKT: 414 AA.
A* Q827P1:
D 01-MAR-2002 (TREMblrel, 20, created)
D 01-MAR-2002 (TREMblrel, 20, last sequence update)
D 01-JUN-2002 (TREMblrel, 21, last annotation update)
DE Glucose-1-phosphatase (G1Pase), secreted.
GN SPY1153.
OS Salmonella typhi.
OW Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OW Salmonella.
OX NCBI_taxid 601;
RR SEQUENCE FROM N.A.
RE STRAIN CT18;
RX MEDLINE 21544447; PubMed-11677608;
RA Parkehill J, Isompan K, Loo K, Bontley S, D., Holden M, T.G., Sobahia M.,
RA Churher C., Munnali K., Brooks K., Chellidharach T., Conerton P.,
RA Baker S., Bueham P., Davies R.M., Dowd L., White N., Farrar J.,
RA Cronin A., Davis P., Hamlin N., Harper A., Hien T.T., Holroyd S., Jansis K.,
RA Kiehl A., Latson T.S., Leather S., Menle S., O'Gaora P., Parry C.,
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrell B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18."
RL Nature 413:848-852(2001);
DR EMBL: AL627269; CAP08242.1;
DR InterPro: IPR00560; HisAc-phosphatase.
DR Pfam: PF00428; acid_phosphatase_1.
DR PROSITE: PS00616; HIS_ACID_PHOSPHAT_1; 1.
DR PROSITE: PS00778; HIS_ACID_PHOSPHAT_2; 1.
KW Complete proteome.
SQ SEQUENCE 414 AA: 45630 MW: 94956.34RRRQNP26 CRC64;

Alignment Scores:
Pred. No.: 2,36e-48 Length: 413
Score: 547.00 Matches: 138
Percent Similarity: 49.88% Conservative: 6
Best Local Similarity: 33.74% Mismatches: 181
Query Match: 23.69% Indels: 24
DB: 16 Gaps: 10

US-09-866-379a-9 (1-1308) x Q827P1 (1-413)

QY 64 GTCACAGTCAGCGGAG---CTCAAGTCGGAAGTCGTGATTCCTACGTCGTCAGT 120
    |||||
DB 22 AlaGlnThrThrProGluGlyTyrGlnLeuGlnGlnValLeuMetMetSerArqHisAsn 41
    |||||
QY 121 GTCAGTCGTCACA---ACCAAGAGCAAGCAACTGATGATGATGATGATGATGATGATGATG 177
    |||||
DB 42 LeuArASuAlaProLeuAlaAsuAsuGlyAsuValLeuAlaGlnSerThrProAsuAlaTrp 61
    |||||
QY 178 GAAATCGAGCGTAAACATGCTGACGCTGACACCGCGCGCTGCTGACCTAATCGCTAT 247
    |||||
DB 62 ProAlaTrpAspValProGlyGlnIleSerThrLysGlyValLeuGluValTyr 81
    |||||
QY 238 CTGAGAGATATACGGCTAGCGTCTGTAGCGGAGATTCGTGGCTAAATGGCGTGC 297
    |||||

```

RESULT 11

Q9AA04

ID

PRELIMINARY: PKT: 414 AA.

```

DB 82 MetGlyPheSerThrArqGlnTrpLeuValAlaGlnGlyLeuThrProSerGlyLeuCys 101
    |||||
QY 298 CAGCAGCTCGGTAGCTCGCGGATTATTTGATGCTGCAAGCAAGCAAGCAAGCAAGCAAGCA 417
    |||||
DB 102 ProAlaTrpAspThrValTyrAlaTyrAlaAsuSerLeuThrValAlaTrpAla 123
    |||||
QY 358 GAAGCCTCGCGCGCGCTGGCACTGATCTGCTGCAATAAACGTCATCAATCAATCAATCAAT 417
    |||||
DB 122 GlnPheIleThrSerAlaPheProGlyCysAspThrProValHisHisGlnGlnGln 141
    |||||
QY 418 AGTCGAGTCGCGGATGCTTATTAATCTCTAAACATGCGGCTGCTGCTGCTGCTGCTGCTG 477
    |||||
DB 142 MetGlyThrMetAspProThrPheAsuProValThrAspAspGlnAlaAlaPheArq 161
    |||||
QY 478 GCGAATCTGATCACTAGCGGATCTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 177
    |||||
DB 162 GlnGlnAlaValGlnAla---MetGlnLysAlaTrpAsuSerGlnGln 196
    |||||
QY 538 CATTATCAAAACGCGCTTCGCGAATCGCAACGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 196
    |||||
DB 176 HisLeuAspGluSerTyrLysLeuLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 196
    |||||
QY 598 TGCGTTAAAGCGTCAGAAACAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 417
    |||||
DB 196 Cys-----LysGlnLysHisGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 411
    |||||
QY 658 CTCAAAGCTCAACGCGGCTGCTCTCTATTAATCAATCAATCAATCAATCAATCAATCAAT 177
    |||||
DB 212 SerAlaAsuTyrGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 196
    |||||
QY 718 AGCAGATATATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 171
    |||||
DB 222 ValAspAlaPheThrLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 196
    |||||
QY 772 AGATACGCTGATTCATCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 196
    |||||
DB 252 GlyLeuHisThrAspArqIleThrLysValLeuSerLysLeuLysAsuSerTyrThrAsp 171
    |||||
QY 832 TTGCTCAACGCGCGGAGGTGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 196
    |||||
DB 272 SerLeuPheThrSerProThrValAlaTrpAsuValAlaAlaAlaProLeuValLysTyr 196
    |||||
QY 892 AAGACAGCGTTGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 196
    |||||
DB 292 AspLysValLeuVal-----AlaGlnAlaValSerAlaPro 196
    |||||
QY 952 TCAGTCGCTGTTATGCGCGGACACGATCTAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1011
    |||||
DB 304 LysValThrValLeuValGlyHisAspSerAsuThrAlaSerLeuLeuThrAlaLeuAsp 424
    |||||
QY 1012 CTCAAC---TGGAAGCTTCGCGGCTGACGCGCGGATCAACAGCGCGGAGCGGCGGAGTGG 1008
    |||||
DB 324 PheLysThrGlyThrLeuHisAspGlnValGlnArqThrProIleGlyGlyGlnVal 196
    |||||
QY 1069 TTTCAACGCTGCGGCTCGGCTAAGCGGATAACAGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1178
    |||||
DB 344 PheGlnArqTrpHisAspGlyAsuAlaAsuArqAspLeuMetLysIleGlyValTyr 196
    |||||
QY 1129 CAGACTTACAGCAGATGCTGATGATAAAACGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1188
    |||||
DB 364 GlnSerAlaArqGlnLeuArqAsuAlaGlnAlaLeuThrLysSerProAlaTrpArq 483
    |||||
QY 1189 GTCAAACTGACCTCGGAGGATGAGAAAGAGGAAAGAGGAGGAGGAGGAGGAGGAGGAGGAG 1248
    |||||
DB 484 ValThrLeuGlnLeuLysGlyCysProVal AspAlaAsuThrPheCysProLeuAsp 407
    |||||
QY 1249 GATTTCAGCAAAACGCTGGAATGAAGCA 1275
    |||||
DB 403 LysPheAspAsuValMetAsuThrAla 411
    |||||
RESULT 11
Q9AA04
ID Q9AA04 PRELIMINARY: PKT: 414 AA.

```

AC Q9AAQ4;  
 DT 01-JUN-2001 (TREMBLrel. 17, Created)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Periplasmic phosphoanhydride phosphohydrolase.  
 GN C0542;  
 OS Caulobacter crescentus.  
 OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;  
 OC Caulobacter;  
 OX NCBI\_TaxID=155892;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC SURAIN-ATCC 19089 / CB15;  
 PX MEDLINE=21172648; PubMed=11259647;  
 RA Norman W.C., Feldhuyzen I.V., Taub M.T., Paulsen I.T., Nelson K.E.,  
 FA Felsen I., Hild-Brandt T.F., Alvey M.F.F., Chiu N., Mullick T.P.,  
 RA Potocka L., Nelson M.C., Newton A., Stephens C., Phadke N.D., Ely B.,  
 RA Deboy K.F., Jackson K.J., Durkin A.S., Gwinn M.L., Hatt D.H.,  
 PA Kolesch J.F., Smit J., Graves M.R., Khouri H., Spott J., Berry K.,  
 RA Osterback T., Tran K., Wolf A., Varadarajan I., Emelianova M., White O.,  
 RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;  
 RT "Complete genome sequence of Caulobacter crescentus";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001)  
 DE D84L, AE095727, AAK22223.1;  
 DR HSSP; P07102; 10KL;  
 DR HCR; C0542;  
 DR InterPro; IPR000560; HisAc\_phosphatase.  
 DR Pfam; PF00328; acid\_phosphatase\_1;  
 DR PROSITE; PS00616; HIS\_ACID\_PHOSPHAT\_1; UNKNOWN\_1.  
 KW Hydro-lase; complete proteome.  
 SQ SEQUENCE 414 AA; 44279 MW; 71050245A01C4051 CRC64;

## Alignment Scores:

Pred. No.: 2,076-37  
 Score: 536.00  
 Length: 414  
 Matches: 145  
 Percent Similarity: 48.03%  
 Conservat: 62  
 Mismatches: 186  
 Best local Similarity: 33.64%  
 Query Match: 23.21%  
 Indels: 38  
 Gaps: 11

us-09-866-379a-9 (1-1308) x Q9AAQ4 (1-414)

QY 25 TTAATCTCTT-----CTATTCCGTTAAACCCCGCAATCTGATTCGTCAGAGTGAG 75  
 DB 3 LeuSerLeuArgHicAlaValValThrLeuAlaLeuAlaSerAlaGlyAlaAlaSerAla 22  
 QY 76 GCGAAGTGGAGCTGAAGAAATGTGTATTCCTACTGATGAAAGTGTGAGTGTGAGT 135  
 DB 23 Glu-----ThrLeuGluLysValValIleLeuSerArgHisGlyValArgSerAlaMet 40  
 QY 136 AAGGGAAGGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 195  
 DB 41 SerSerProGluAlaLeuGluAlaSerAlaAlaProIleProIleGluValPro 60  
 QY 196 CTGGCTGAGTGTACACCG 255  
 DB 61 AlacIlyHisLeuThrAlaValGlyValThrLeuValAlaAlaArgMetGlyAspTyrArg 80  
 QY 256 CAGGCTGTGTGAGCGAGGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 315  
 DB 81 ArgHisTyrAlaAlaGlyValLeuLeuLysProTyrAspGly-----AlaSerVal 97  
 QY 416 GATATATCTCTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 375  
 DB 98 TyrAlaTrpAlaAsnValThrGluArgThrIleAlaThrAlaLysAlaTyrArgGluThr 117  
 QY 476 CTGGCACTGATCTGCTGCAATACCGGTACAlaTCCAGGCGAGAlaAGTCCAGTCCGATCG 435  
 DB 118 LeuAlaProGlyGlySerValThrValAsnThrValGlyGlu---GlyAsuIleAspPro 136  
 QY 436 TTAATTAATCTCTAAAACCTGGGCTTTGCTAAATGTGATGATGATGATGATGATGATG 495  
 DB 137 MetIleGluProValIlyAlaGlyIleValIlyAlaAspHisAlaLeuAlaArgAlaAla 156

QY 496 ATCTCTCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 555  
 DB 157 ValAlaGlyValGlyGlyAspLeuThrAlaValSerHisAsnGluAla 176  
 QY 556 GCGAAGTGGAGCTGAAGAAATGTGTATTCCTACTGATGAAAGTGTGAGTGTGAGT 615  
 DB 177 GluGluGluAspAlaLeuGluGluGlyCysAspGlyGlyProGlyGlyGlyGlyGly 191  
 QY 616 CAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 666  
 DB 192 -----ProProAlaProGlyLysAsnValIleAspAla 203  
 QY 667 -----AGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 708  
 DB 204 TysProGlyPheValAspGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 223  
 QY 709 TCAATGCGAGAGGAGATATTTCTGCGCAACAAAGAAAGAGGAGGAGGAGGAGGAGGAG 768  
 DB 224 SerGlyValThrIleSerLeuLeuMetAlaTrpAlaAspGlyArgAspPheAlaGlyLeu 243  
 QY 769 GGA---AGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 825  
 DB 244 GlyTrpSerLeuAspGlyAlaLeuThrArgSerPheLeuHisSerAlaGly 263  
 QY 826 TTTGATTTGCGTACACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 885  
 DB 264 PheAspLeuAlaLeuArgThrProTyrValAlaAlaThrLeuAlaGlyHisLeuAlaAsp 283  
 QY 886 TTTGATTTGCGTACACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 945  
 DB 284 ArgLeuAlaAlaThrLeuArg-----AspGlyAlaAlaAlaIleGly---ProVal 299  
 QY 946 CCCACTTCAGTCTCTTTTATCCCGGACAGATACATACTGCGCAATCTAGGAGGAGGAG 1005  
 DB 300 AspAlaArgLeuValIleAlaGlyHisAspGlyThrLeuAlaSerLeuGlyGlyLeu 419  
 QY 1006 CCGCAGCTCACTGCGCGCTTCCCGCT---CAGGCGGATAAACCGCGCGGAGGAGGAG 1062  
 DB 320 LeuArgMetGlyThrThrLeuProGlyTyrGlyPheAsnGlyIleLeuProGlyGlyAla 339  
 QY 1063 CTTGCTCTTGAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1122  
 DB 340 LeuValPheGluArgTrpArgArg---AspAspGlyValArgValValArgValArgPhe 358  
 QY 1123 GTCCTCTACACCTTTACAGCAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1182  
 DB 359 ThrGlyGlnSerLeuSerGlnLeuArgAsnMetThrAlaLeuAspAlaLysThrProPro 378  
 QY 1183 GAGAGGCTGAAACTGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1242  
 DB 379 LeuSerAlaProValPheValGlyGlyGlyGlyThrAlaThrProAlaPheAspCysArg 498  
 QY 1243 TTGGAGCTTTTATCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1275  
 DB 399 LeuGluAspPheGluThrValValArgGlyAla 409  
 RESULT 12  
 Q8VUS2 PRELIMINARY; PRI: 419 AA.  
 ID Q8VUS2  
 AC Q8VUS2;  
 DT 01-MAR-2002 (TREMBLrel. 20, Created)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 21, Last annotation update)  
 DE Phytase.  
 GN PHY.  
 OS Klebsiella pneumoniae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Klebsiella;  
 OX NCBI\_TaxID=573;  
 RN [1]  
 RP SEQUENCE FROM N A  
 RC STRAIN-ASR1;







```
QY 439 TTAAATGCTCTAAAAATGGGGTTTTCGCAACCGGATAACCGAAGCGGACGTCGACGCGGATC 498
Db 95 ----- 95
QY 499 CTCGAGAGGCGACGAGAGGTGCAATTGCTGACTTTACCGGGATATTATTAATAATGAGGTTCGC 558
Db 96 ----- 101
QY 559 GAATCGAAGCGGTGGCTTAATTTTCGCGCAATCAAACTGGTGGCTTAAACCTGACAAACAG 618
Db 101 ----- 101
QY 619 GAGGAAAGCGCTTCATTAAAGGAGGAGGATTAACCATGGAATCAAGGTGACGGCGGACTGT 678
Db 101 ----- 101
QY 679 GTCTCATTACCGGT---GGGATAAGGCTCGCATCAATGCTGACGAGAGATATTTCCTCG 735
Db 102 LeuSerLeuValGlyLeuPheLeuArgAlaIlePheValLeuAspThrLeuMetCys 121
QY 736 CAACAGACAGAGGGAATCGCGGAGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 795
Db 122 GluAlaIleHisGlyLeuProLeuProAlaIlePheAla-----SerPro 135
QY 796 AATACCTTGCTAAGTTTCGATACAGCGGCAATTTCATTTGCTGATACAGCGGAGGAGGAGGAG 855
Db 136 AspValLeuArgThrLeu-----AlaGlnIleSerAlaLeuAspIleGlyAlaIleVal 153
QY 856 GCGGAGGAGCGCGGCGGCGGCTTATTATGTTG-----ATCAACAGACAGCGTTG 903
Db 154 GlyProArgAlaAlaGlnIleValAlaGlnLeuThrGlyGlyIleLeuLeuAsnAlaIle 173
QY 904 AGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 963
Db 174 LeuAlaAsnPheSerArgValGlnArgLeuGly-----LeuProLeuLysMetValMet 191
QY 964 ATCGCGGAGACGATATCTAATCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1023
Db 192 TyrSerAlaIleHisAspSerThrLeuLeuAlaLeuGlnGlyAlaLeuGlyLeu----- 208
QY 1024 GTTCGCGGCTAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1080
Db 209 TyrAspGlyHisThrProProTyrAlaAlaCysLeuGlyPheGluPheArg 225
QY 1081 GGTGCGGCTAAGCGGATACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1131
Db 226 CysHisLeuGlyAsnProAlaLysAspGlyGlyAsnValThrValSerLeuPheTyrArg 245
QY 1132 ACTTTACAGCAGATGGGTGATAAAGCGCGGCTGCTGCTTAAATACGCGGAGGAGGAGGAGGAG 1191
Db 246 AsnAspSerAlaIleHisLeuProLeu----- 253
QY 1192 AAATGAGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1251
Db 254 PheLeuSerLeuGlyGlyCys-----TTCAlaIleProCysProLeuGlyArg 268
QY 1252 TTATGAGCAAAATGCGTAATGAAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1305
Db 269 PheTyrGlnLeuThrAlaProAlaAsnProProAlaHisGlyValSerCysHis 286
```

Search completed: December 13, 2002, 16:05:13  
Job time : 105 secs







```

Db 21 PheAlaInSerGluProGluLeuLysLeuGlnSerValValIleValSerArqHisGly 40
QY 121 GGGAGATGTCACAGCAAGAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 180
Db 41 ValArgAlaProThrLysAlaThrGlnLeuMetGlnAspValThrProAspAlaTrpPro 60
QY 181 ACCTGGAGGTAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 240
Db 61 ThrTrpProValLysLeuGlyTrpLeuThrProArgGlyGlyGlnLeuIleAlaLysLeu 80
QY 241 GACATTACTGGCTGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 300
Db 81 GlyHisTrpGlnAspGlnAspGlnAspGlnAspGlnAspGlnAspGlnAspGlnAspGln 100
QY 301 CAGCTGGCTGACGCTGACGCTGACGCTGACGCTGACGCTGACGCTGACGCTGACGCTG 360
Db 101 GlnSerGlyGlnValAlaIleIleAlaAspValAspGlnArgTrpArgLysThrGlyGln 120
QY 361 GACTTCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 420
Db 121 AlaPheAlaAlaGlyLeuAlaProAspCysAlaIleThrValHisThrGlnAlaAspThr 140
QY 421 TCGACTCGGATCGGCTATTATTAATCTCTTAAAGCTGGCTTTGGCAACTGGATAACGG 480
Db 141 SerSerProAspProLeuPheAsnProLeuLysThrGlyValCysGlnLeuAspAsnAla 160
QY 481 AATCTGATCTACGGATCTGACAGCGACAGCGCTCAATTCGTGACTTTACCGGCAT 540
Db 161 AsnValThrAspAlaIleLeuSerArgAlaGlyGlySerIleAlaAspPheThrGlyHis 180
QY 541 TATTAAGAGGCTTTCGGAAATGGAACGGGTGTTAATTTTCGGCAATCAAACTTTGCG 600
Db 181 ArgGlnThrAlaPheArgGlnLeuGlnArgValLeuAsnPheProGlnSerAsnLeuCys 200
QY 601 CTAAAGATGAGAAACAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 660
Db 201 LeuLysArgGlnLysGlnAspLeuSerCysSerLeuThrGlnAlaLeuProSerGlnLeu 220
QY 661 AAGGAGAGCGCGGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 720
Db 221 LysValSerAlaAspAsnValSerLeuThrGlyAlaValSerLeuAlaSerMetLeuThr 240
QY 721 GAGATATTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 780
Db 241 GlnIlePheLeuLeuGlnGlnAlaGlnGlyMetProGlyTrpGlyArgIleThr 260
QY 781 GATTACACCAAGTGGCAACACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840
Db 261 AspSerHisGlnTrpAsnThrLeuLeuSerLeuHisAsnAlaGlnPheTrpLeuLeuGln 280
QY 841 GGTACACCAAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 900
Db 281 ArgThrProGlnValAlaArgSerArgAlaThrProLeuLeuAspLeuIleMetAlaAla 300
QY 901 TTGACGCGCATCCACCGCAAAACAGCGGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 960
Db 301 LeuThrProHisProGlnLysGlnAlaTrpGlyValThrLeuProThrSerValLeu 320
QY 961 TTATGCGCGCACAGATATTAATCTGCAAAATCTGCGCGCGCGCGCGCGCGCGCGCGCG 1020
Db 321 PheIleAlaGlyHisAspThrAsnLeuAlaAsnLeuGlyGlyAlaLeuGlnLeuAsnTrp 340
QY 1021 ACATTCTCGGCTGACGCGGATAAAACGCGCGCAATCTGCTGCTGCTGCTGCTGCTGCTG 1080
Db 341 ThrLeuProGlyGlnProAspAsnThrProGlyGlyGlnLeuValPheGlnArgTrp 360
QY 1081 GCGGCGCAACCGCATACGCGCATCGGATTCAGGCTTTCGCTGCTGCTGCTGCTGCTGCTG 1140
Db 361 ArgArgLeuSerAspAsnSerGlnTrpIleGlnValSerLeuValPheGlnThrLeuGln 380
QY 1141 CAAATGCTGATATAAAAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1200
Db 381 GlnMetArgAspLysThrProLeuSerLeuAsnThrProGlyGlnValLysLeuThr 400

```

```

QY 1201 CTGGAGATGTCACAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 1200
Db 401 LeuAlaLysLysGlnArgAsnAlaLysLysLysLysLysLysLysLysLysLysLysLys 4200
QY 1261 AACTGGATGACAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 1300
Db 421 IleValAsnGlnAlaArgAlaProAlaArgSerLeuAlaArgSerHis 4400
RESULT 2
US-09-318-528-2
; Sequence 2, Application: us/09-318-528
; Patent No. 6194740
; GENERAL INFORMATION:
; APPLICANT: Kretz, Keith
; TITLE OF INVENTION: NOVEL PHYTASE
; FILE REFERENCE: 09010/029004
; CURRENT APPLICATION NUMBER: US/09-318-528
; CURRENT FILING DATE: 1999-05-25
; EARLIER APPLICATION NUMBER: 09/291,931
; EARLIER FILING DATE: 1999-04-13
; EARLIER APPLICATION NUMBER: 08/910,798
; EARLIER FILING DATE: 1997-08-13
; EARLIER APPLICATION NUMBER: 09/259,214
; EARLIER FILING DATE: 1999-03-01
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 440
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-318-528-2
Alignment Scores:
Pred. No.: 3,530-221 Length: 440
Score: 2188.00 Matches: 425
Percent Similarity: 97.70% Conserved: 10
Best Local Similarity: 97.70% Mismatches: 10
Query Match: 94.76% Indels: 0
DB: 4 Gaps: 0
US-09-866-379a-9 (1-1308) x US-09-318-528-2 (1-440)
QY 1 AIGAAAGCATCTTAATCTGATTTTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 40
Db 1 MetLysAlaIleLeuIleProPheLeuSerLeuLeuIleProLeuThrProLeuSerAla 20
QY 61 TTGCTTCAGCTGACGCTGACGCTGACGCTGACGCTGACGCTGACGCTGACGCTGACGCTG 120
Db 21 PheAlaGlnSerGluProGluLeuLysLeuGlnSerValIleValSerArqHisGly 40
QY 121 GGGAGTGTCCCAACCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 180
Db 41 ValArgAlaProThrLysAlaThrGlnLeuMetGlnAspValThrProAspAlaTrpPro 60
QY 181 ACCTGGAGGTAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 240
Db 61 ThrTrpProValLysLeuGlyTrpLeuThrProArgGlyGlyGlnLeuIleAlaLysLeu 80
QY 241 GACATTACTGGCTGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 300
Db 81 GlyHisTrpGlnArgGlnArgGlnValAlaAspGlyGlnLeuAlaLysLysLysLysLys 320
QY 301 CAGCTGGCTGACGCTGACGCTGACGCTGACGCTGACGCTGACGCTGACGCTGACGCTG 360
Db 101 GlnSerGlyGlnValAlaIleIleAlaAspValAspGlnArgTrpArgLysThrGlyGln 420
QY 361 GACTTCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 420
Db 121 AlaPheAlaAlaGlyLeuAlaProAspCysAlaIleThrValHisThrGlnAlaAspThr 440
QY 421 TCGACTCGGATCGGCTATTATTAATCTCTTAAAGCTGGCTTTACCGGCAT 480

```



QY 841 CCGACGACAGAGTTCGGCGGACGCGGACCGCGCTATTAGATTGATCAAGACAGCG 900  
|||||  
Db 281 ArqThrProGluValAlaArqSerArqAlaThrProLeuLeuAspLeuLeuMetAlaAla 300  
|||||  
QY 901 TTACAGCCGACGACGCGGAAAAACAGAGGTATGAGTGTGACATTACGACCATTTACGTGTC 460  
|||||  
Db 301 LeuThrProHisProGlnGlySerLeuAlaTyrGlyValThrLeuProThrSerValLeu 320  
|||||  
QY 961 TTTATCGCGGCAAT 1020  
|||||  
Db 321 PheIleAlaGlyHisAspThrAsnLeuAlaAsnLeuGlyGlyAlaLeuGluLeuAsnTrp 340  
|||||  
QY 1021 ACATTTCTGGGTGAGGCGGATAAATATATATATATATATATATATATATATATATATAT 1080  
|||||  
Db 341 ThrLeuProGlyGlnProAspAsnThrProProGlyGlyGluLeuValPheGluArgTrp 360  
|||||  
QY 1081 CGTCGGGTAACTAT 1140  
|||||  
Db 361 ArgArgLeuSerAspAsnSerGlnTrpIleGlnValSerLeuValPheGlnThrLeuGln 380  
|||||  
QY 1141 CAGATCGGTGAT 1200  
|||||  
Db 381 GluMetArgAspLysThrProLeuSerLeuAsnThrProProGlyGlyGluValLysLeuThr 400  
|||||  
QY 1201 CTCTAT 1260  
|||||  
Db 401 LeuAlaGlyTyrGlnIleAlaAsnAlaGlnGlyMetCysSerLeuAlaGlyPheThrGln 420  
|||||  
QY 1261 AT 1320  
|||||  
Db 421 IleValAsnGluAlaArgIleProAlaCysSerLeuArgSerHis 440  
|||||

## RESULT 4

US-08-910-798-2  
Sequence 2, Application US/08910798  
Patent No. 5876997  
GENERAL INFORMATION:  
APPLICANT: KRETZ  
TITLE OF INVENTION: NOVEL PHYTASE  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 4225 Executive Square, Suite 1400  
CITY: La Jolla  
STATE: California  
COUNTRY: US  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/910,798  
FILING DATE: August 13, 1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: HAILE, PH.D., LISA A.  
REGISTRATION NUMBER: 38,347  
REFERENCE/DOCKET NUMBER: 09010/029001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619/678-5070  
TELEFAX: 619/678-5099  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 423 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-910-798-2

Alignment Scores:  
Prod. No.: 6,690-210 Length: 423

Score: 2080.50 Matches: 408  
Percent Similarity: 93.79% Conserved: 6  
Best Local Similarity: 93.79% Mismatches: 10  
Query Match: 90.10% Indels: 17  
DB: 2 Gaps: 1  
US 09 866 379A 3 (1-1308) x 05 08 910 798 2 (1-423)  
QY 1 AUGAAGAGAGUATTAATCGGCAATTTTAACTGCTCTGAGATTCGCTGAAAGGCGAAAGTAA 60  
|||||  
Db 1 MetLysAlaIleLeuIleProPheLeuSerLeuLeuIleProGluHisIleLeuSerAla 20  
|||||  
QY 61 TTTCTTCAGATGAGAGTGGAGCTCAAGCTGCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 120  
|||||  
Db 21 PheAlaGlnSerGluProGluLeuLysLeuGluSerValValIleValSerAlaHisSly 40  
|||||  
QY 121 GGTGAGTGTGAAACAAAGAGGCAAGCAATGATGACAGATGAGTACGAGAGAGAGAGAGAGAG 180  
|||||  
Db 41 ValArgAlaProThrLysAlaThrGlnLeuMetGlnAspValThrProAspAlaTrpPro 60  
|||||  
QY 181 AATGAGAGAGTAAAGATGAGTGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240  
|||||  
Db 61 ThrTrpProValLysLeuGlyTrpLeuThrProGlyLysLeuAlaGlyLeuAlaLysLeu 80  
|||||  
QY 241 GGCATTTACTGGGAGTCAAGAGTCTGAGTACAGGAGAGAGATGAGTGAAGAGAGAGAGAGAG 300  
|||||  
Db 81 GlyHisTyrGlnAlaGlnArgLeuValAlaAspGlyLeuLeuAlaLysLeuGlyLysPro 100  
|||||  
QY 301 CAGTCTGCTCAGTCTGAGATTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360  
|||||  
Db 101 GlnSerGlyGlnValAlaIleIleAlaAspValAspIleAlaTrpArgLysThrLys 119  
|||||  
QY 361 GCGTTTCAGCGCGGAGCTGAGACGCTGAGTGTGAGTGAATGAATGAATGAATGAATGAATGA 420  
|||||  
Db 120 ----- GlnAlaAspThr 124  
|||||  
QY 421 TCCAGTCGCGCATCGTTATTTAAATCGCTCAAAAATGCGCTTTCTGCACTGATGAACGCG 480  
|||||  
Db 124 SerSerProAspProLeuPheAsnProLeuLysThrGlyValLysGlnLeuAspAla 144  
|||||  
QY 481 AAGTGACTGACGGATCTCGACAGAGGACAGAGAGGCTCAATTTGCTGATTTTAAAGAGAT 540  
|||||  
Db 144 AsnValThrAspAlaIleLeuSerArgAlaGlyLysSerIleAlaAspPheThrLysHis 164  
|||||  
QY 541 TATCAAAAGCGGTTTCAGAACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600  
|||||  
Db 164 ArgGlnThrAlaPheArgGluLeuGluLeuValLeuAsnProIleSerAsnLeuGly 184  
|||||  
QY 601 CTTAAAGTGTGAGAAACAGACGCGAAAGTGTGCTATTAAAGAGGATTTACTATACGAAATP 660  
|||||  
Db 184 LeuLysArgGluLysGlnAspGluSerCysSerLeuThrGlnAlaLeuProSerGlnLeu 704  
|||||  
QY 661 AAGCTGAG 720  
|||||  
Db 204 LysValSerAlaAspAsnValSerLeuThrGlyAlaValSerLeuAlaSerMetLeuThr 724  
|||||  
QY 721 GAGATATTTCTCCCTGCAACAAAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780  
|||||  
Db 224 GluIleThrLeuLeuGlnIleAlaGlnGlyMetProGlnArgGlyLysTrpAlaAlaLeu 800  
|||||  
QY 781 GATTCAACACGAGTGGAAACACCTTGTGTAAGTTGTGATTAAGAGAGAGAGAGAGAGAGAGAG 840  
|||||  
Db 244 AspSerHisGlnTrpAsnThrLeuLeuSerLeuHisAsnAlaIlePheTyrLeuLeuGln 864  
|||||  
QY 841 GCGACGCGCAGAGGTTGGCGCGAGCGCGCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900  
|||||  
Db 264 ArgThrProGluValAlaArgSerArgAlaThrProLeuLeuAspLeuIleMetAlaAla 924  
|||||  
QY 901 TTGAG 960  
|||||  
Db 294 LeuThrProHisProGlnLysAlaTyrGlyValIleThrProSerIleValLeu 1004  
|||||  
QY 961 TTTATCGCGCGACACGATATTAATTCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020  
|||||





















```

QY 817 -----AAGCGCAATTTGATTTGCTACAAAGCGACGACAGGTTGGCGCGAGC 864
DB 311 AlaValTyrAlaAsnAlaSerLeuGluAsnGlnGlyProGlySerGluAla----- 328
QY 865 GAGGACAGCGCGTTATTATGATTCATCAACAAAGATCTTCACGCGCATCCACCGCAAAAA 924
DB 328 ----- 328
QY 925 CAGCGTATGCTCTACATTACGACATTCAGCTGCTTTATAGCGCGACACAGGATACAAAT 984
DB 329 -----GlySerLeuPhePheAsnPheAlaHisAspThrAsn 340
QY 985 CTGCGAAATCTGCGCGCGCTGACTGAGTCTG-----AATCGACAGCTTCTCC----- 1029
DB 341 HisThrProIleLeuAlaAlaLeuGlyValLeuIleProAsnGluAspLeuProLeuAsp 360
QY 1030 -----GGTCAGCGCG-----GATAACACGCGCGCGAGCTGCTGAACTG 1065
DB 361 ArgValAlaPheGlyAsnProTyrSerIleGlyAsnIleValProMetGlyGlyHisLeu 380
QY 1066 GTCTTGACCGCTGG-----CGTCGATTAAGCATATACACGACGACGCTGATTCAG 1113
DB 381 ThrIleGluAlaLeuSerCysGlnAlaThrAlaLeuSerAspGlnGlyThrTyrValArg 400
QY 1114 GTTCTGCTGCTCTTCAGACTTTACAGCATGCTGTGATAAAACGCGGCTGCTCATTAAAT 1173
DB 401 LeuValLeu----- 403
QY 1174 AGCGCGCGCGACAGCTTGAACGTCAGCGCGCGCGACAGCGCGCGCGCGCGCGCGCGCG 1223
DB 404 -----AsnGluAlaValLeuProPheAspAspCysThrSerGlyProGlyTyrSer 420
QY 1244 AATCTGCTGCTGCTGCTTTCAGCTTTCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1272
DB 421 CysProLeuAlaAsnSerThrSerIleLeuAsnIle 432
RESULT 14
US-08-609-426A-2
Sequence 2, Application US/98/09426A
Patent No. 5840733
GENERAL INFORMATION:
APPLICANT: Nevalainen, Helena K.M.
APPLICANT: Paloheimo, Maria T.
APPLICANT: Miettinen-Oinonen, Arja S.K.
APPLICANT: Torkkeli, Tuula K.
APPLICANT: Cantrell, Michael
APPLICANT: Piddington, Christopher S.
APPLICANT: Ramesek, John A.
APPLICANT: Turunen, Marja K.
APPLICANT: Paquerström, Richard B.
APPLICANT: Houston, Christine S.
TITLE OF INVENTION: Production of Phytase Degrading Enzymes
NUMBER OF SEQUENCES: 69
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox
STREET: 1100 New York Avenue, Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/98/09426A
FILING DATE: 01-MAR-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/923,724
FILING DATE: 31-JUL-1992

```

```

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/496,155
FILING DATE: 19-MAR-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/044,977
FILING DATE: 29-APR-1987
PRIOR APPLICATION DATA:
APPLICATION NUMBER: UK 9610600
FILING DATE: 30-APR-1986
ATTORNEY/AGENT INFORMATION:
NAME: Reed, Grant E.
REGISTRATION NUMBER: P-41,264
REFERENCE/POCKET NUMBER: 1950,0080001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 479 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-609-426A-2
Alignment Scores:
Pred. No.: 0.00625 Length: 479
Score: 109.50 Matches: 91
Percent Similarity: 29.49% Conservatives: 48
Best Local Similarity: 19.21% Mismatches: 145
Query Match: 4.74% Indels: 189
Gaps: 24
US 09 866 379A 9 (1 1208) x 95 98 409 426A 2 (1 479)
QY 85 AAGCTGAAAGATGAGGATGCTACAGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 144
DB 72 GluValAspGlnValIleMetValIysAlaHisGlyGluAlaGlyProGlySerProSerAla 91
QY 145 CAACGAGACAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 190
DB 92 Gly LysAspIleGluAlaLeuAlaLeuAlaLeuAlaLeuAlaLeuAlaLeuAlaLeuAla 109
QY 199 -----GGTAGAGTGACA----- 210
DB 110 TyrLysGlyAspLeuAlaPheLeuAsnAspIleThrTyrValThrAsnLeuLysTyr 129
QY 211 -----CGCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 242
DB 130 TyrAsnAlaGluThrThrSerGlyProTyrAlaLeuAlaAspAlaTyrAsnHisLys 149
QY 238 -----CTGGGAGATATAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 279
DB 150 AsnAspTyrLysAlaLysTyrGlyHisLeuIlePheGlyValLeuValValThrPhePhe 169
QY 286 CTGGCTAAATGTGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 409
DB 170 -----SerSerGlyTyrGlyAlaValIle 178
QY 340 GGTACCGGTAAACAGGCGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 496
DB 179 ThrAlaAspGlySerGlyGlyGlyPheGlyPheGlyGlyGlyGlyGlyGlyGlyGlyGly 496
QY 397 ACCGTACATACCGGACAGATACGCTGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 496
DB 199 AsnIleThrSerGluSerGluValMetGlyAlaAspSerLeu 219
QY 457 GAGGCTTGGCAACCTGGATTAACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 516
DB 214 ProThrLysAspThrAspAsnAspIleThrThrLysAspAsnLeuThr 229
QY 517 TCAATTGCTGACTTACCGGCGCATATATCAAAAGCGGCTTTCAGGAGGAGGAGGAGGAGGAG 576
DB 230 -----TyrGln----- 241

```







denCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model  
Run on: December 13, 2002, 15:55:21, Search time 62 Seconds  
(without alignments)  
5622.316 Million cell updates/sec  
Title: US-09-866-379A-9  
Perfect score: 2009  
Sequence: 1 atgaagcgatcttaatcc..... gaggatggagctctatcta 1308

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 1816940

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame\_n2p.model -DEV=xlp  
-Q=us-09-866-379a-9 -FASTA=us-09-866-379a-9.fasta\_1 1479  
-DB=A\_Geneseq\_101002 -OFMT=fastan -SUFFIX=trax -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -STAR=1 -END=1 -MATRIX=BLOSUM62 -TRANS=human40.cd:  
-LIST=45 -DOCATALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -GTFM=pco -NORM=ext -HEAFISE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09866379A -CGN\_1\_1\_45 -runat\_1312052\_14455\_21075 -NCT=6 -ICPU=3  
-NO\_XLPXY -NO\_MAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG DEV\_TIMEOUT=126  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: A\_Geneseq\_101002:  
1: Escherichia coli AAE15808  
2: Escherichia coli AAE15808  
3: Escherichia coli AAE15808  
4: Escherichia coli AAE15808  
5: Escherichia coli AAE15808  
6: Escherichia coli AAE15808  
7: Escherichia coli AAE15808  
8: Escherichia coli AAE15808  
9: Escherichia coli AAE15808  
10: Escherichia coli AAE15808  
11: Escherichia coli AAE15808  
12: Escherichia coli AAE15808  
13: Escherichia coli AAE15808  
14: Escherichia coli AAE15808  
15: Escherichia coli AAE15808  
16: Escherichia coli AAE15808  
17: Escherichia coli AAE15808  
18: Escherichia coli AAE15808  
19: Escherichia coli AAE15808  
20: Escherichia coli AAE15808  
21: Escherichia coli AAE15808  
22: Escherichia coli AAE15808  
23: Escherichia coli AAE15808

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	DB ID	Description
1	2228	96.5	430	23	AAE15808
2	2188	94.8	410	22	Escherichia coli AAE15808
3	2188	94.8	440	23	Escherichia coli AAE15808
4	2188	94.8	440	23	Escherichia coli AAE15808
5	2182	94.5	432	21	AU777755
6	2182	94.5	432	21	AAB36257
7	2182	94.5	432	21	AAB36257
8	2182	94.5	432	21	AAB36257
9	2182	94.5	432	21	AAB36257
10	2182	94.5	432	21	AAB36257
11	2182	94.5	432	21	AAB36257
12	2182	94.5	432	21	AAB36257
13	2182	94.5	432	23	AAB15807
14	2177	94.3	432	22	AAB2631
15	2171	94.0	432	22	AAB2631
16	2170	94.0	432	22	AAB2631
17	2157	93.4	432	22	AAB2631
18	2147	93.0	432	21	AA94753
19	2080.5	90.1	423	20	AA94753
20	1192	51.6	261	22	AB225598
21	676	29.4	140	22	AB225598
22	637	27.6	144	22	AB225598
23	590	25.5	123	22	AB225598
24	324.5	14.1	118	22	AB225598
25	263	11.4	1303	22	AB225598
26	201	8.7	426	22	AB225598
27	201	8.7	426	23	AAU75788
28	201	8.7	426	23	AAU75788
29	184.5	8.0	1062	22	AB225598
30	131.5	5.7	423	22	AAU14067
31	127	5.5	395	22	AB225598
32	126	5.5	440	21	AAV69549
33	126	5.5	465	19	AAW84356
34	126	5.5	465	20	AAV39905
35	125	5.4	440	21	AAU20507
36	125	5.4	440	21	AAU20507
37	125	5.4	467	22	AAU20507
38	124	5.4	467	20	AAV43171
39	124	5.4	467	21	AAU20528
40	124	5.4	467	21	AAV69570
41	124	5.4	467	22	AAU20528
42	124	5.4	467	22	AAU20528
43	124	5.4	467	22	AAU20528
44	118	5.1	440	21	AAU20528
45	118	5.1	440	21	AAU20528

ALIGNMENTS

RESULT 1  
AAE15808  
ID AAE15808 Standard; Protein: 430 AA.  
AC AAE15808;  
DL 29-MAR-2002 (first entry)  
DE Escherichia coli appA phytase mutant protein.  
KW Bacterial phytase, K12 appA phytase, protease, anabolic;  
KW fast-growing, nutritional value, food treatment process; therapy;  
KW chemical tolerance, growth performance; alcoholic drink; biopulping;  
KW non-alcoholic drink, biobleaching; mutant; mutagen.  
OS Escherichia coli.  
US Synthetic.  
XX  
FH Key Location/Qualifiers



```

|||||
Db 379 GlnMetArgAspLysThrProLeuSerLeuAsnThrProProGlyGluValLysLeuThr 398
QY 1261 CTGGACGATGTTCACACCAAAATCGGAGGGTATGCTTCGCTGGGAGCTTTTAAAGCAA 1260
Db 399 LeuAlaGlyCysThrGluValAsnAlaGlnGlyMetCysSerLeuAlaGlyPheThrGln 418
QY 1261 ATGGGGAACCAAGCAATATCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1260
Db 419 IleValAsnGluAlaArgIleProAlaCysSerLeu 430

RESULT 2
AAB47892
ID AAB47892 standard; Protein; 440 AA
XX
AC AAB47892;
DT 07-MAR-2001 (first entry)
DE Escherichia coli B phytase enzyme.
XX
KW Escherichia coli B, phytase enzyme, substrate, substrate, phytate digestion.
XX
OS Escherichia coli.
XX
PN M0200071728-A1.
XX
FO 30-NOV-2000.
XX
PE 25-MAY-2000; 2000WO-US14846.
XX
PR 25-MAY-1999; 9905S-0318528.
XX
PA (DIVE-) DIVERSA CORP.
XX
PI Short JM, Kretz KA;
XX
DR WPI: 2001-112081/12.
XX
DR N-PSDB: AAC88885
XX
PT Improving the nutritional value of phytate-containing feedstuffs, using
PT phytase enzymes which catalyze the liberation of inorganic phosphate
PT from the phytates -
XX
PS Claim 1, Fig 1, 147pp, English.
XX
CC The present sequence is a phytase enzyme from Escherichia coli B.
CC the enzyme catalyzes the liberation of inorganic phosphate from the
CC phytate in phytate-containing feedstuffs and can thus be used to improve
CC the nutritional value of phytate rich ingredients.
XX
SQ Sequence 440 AA;

Alignment Scores:
Pred. No.: 5,966 222 Length: 440
Score: 2188.00 Matches: 425
Percent Similarity: 97.29% Cuser: jcl 0
Best Local Similarity: 97.70% Mismatches: 10
Query Match: 94.76% Indels: 0
DB: 22 Gaps: 0

US-09-866-379A-9 (1-1308) x AAB47892 (1-440)
QY 1 ATCAAGCGATGTTAATGCCATTTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 60
Db 1 MetGluAlaIleLeuLeuProPheLeuSerLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 20
QY 61 TTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 120
Db 21 PheAlaGlnSerGluProGluLeuLysLeuGluSerValIleValIleSerArgHisGly 40
QY 121 GTGGCGGCTGCCAAGCAATGCGACGGCAATGCGACGGCAATGCGACGGCAATGCGACGGCAATGCG 1260

```

```

|||||
Db 41 ValArgAlaPheThrLysAlaIleThrGlnLeuMetGluAspValThrProAspAlaTrpPro 60
QY 181 AATGAGGCTGTAAGATGAGTATGAGTATGAGTATGAGTATGAGTATGAGTATGAGTATGAGTATG 240
Db 61 ThrTrpThrValLysLeuGlyTrpLeuThrPheArgGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 20
QY 241 GAAACATATGAGGCTGAGTATGAGTATGAGTATGAGTATGAGTATGAGTATGAGTATGAGTATG 300
Db 81 GlyHisLysGlnArgGlnArgLeuValAlaAspGlyLeuLeuAlaLysLysGlyCysPro 100
QY 301 CATTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 360
Db 101 GlnSerGlyGlnValAlaIleIleAlaAspValAspGluLeuLeuLeuLeuLeuLeuLeuLeu 120
QY 361 GCGTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 420
Db 121 AlaPheAlaIleGlyLeuAlaProAspCysAlaIleThrValHisThrGluAlaAspThr 140
QY 421 TCCAGTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 480
Db 141 SerSerPheAspPheLeuPheAsnGluLeuLysThrGlyValLysLeuLeuLeuLeuLeuLeu 160
QY 481 AAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 540
Db 161 AsnValThrAspAlaIleLeuSerAlaGlyGlyGlySerIleAlaAspPheThrGlyHis 180
QY 541 TATCAACGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 600
Db 181 ArgGlnThrAlaPheArgGluLeuGluArgValLeuAsnProGlnSerAsnLeuCys 200
QY 601 CTTAAACGCTGAGAACAGCAGCAAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 660
Db 201 LeuLysArgGluLysGluAspGluSerCysSerLeuThrGluAlaLeuProSerGluLeu 220
QY 661 AAGGTGAGGCGGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 720
Db 221 LysValSerAlaAspAsnValSerLeuThrGlyAlaValSerLeuAlaSerMetLeuThr 240
QY 721 GACATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 780
Db 241 GlnIlePheLeuLeuGlnGlnAlaGlnGlyMetGluPheSerProGlyTrpGlyAlaGlyHis 260
QY 781 GATTACATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 840
Db 261 AspSerHisGlnTrpAsnThrLeuLeuSerIleLeuIleAsnAlaGluLeuLeuLeuLeuLeu 280
QY 841 GCGACGCTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 900
Db 281 ArgThrProGluValAlaArgSerArgAlaThrProLeuLeuAspLeuIleMetAlaIle 300
QY 901 TGCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 960
Db 301 LeuThrProHisProGlnLysGlnAlaTyrGlyValThrLeuProThrSerValLeu 420
QY 961 TTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1020
Db 321 PheIleAlaGlnHisAspThrAsnLeuAlaSerLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 340
QY 1021 AGCGTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1080
Db 341 ThrLeuProGlyGlnProAspAsnThrProGlyGlyGlyLeuValPheGluAlaTrp 360
QY 1081 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1140
Db 361 ArgArgLeuSerAspAsnSerGlnTrpIleGlnValSerLeuValPheGlnThrLeuGln 480
QY 1141 CAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1200
Db 381 GlnMetLeuAspLysThrProLeuSerLeuAsnThrProGlyGlyGlyValLysLeuThr 400
QY 1201 CTGGAGGATGCTGAGCAAGCAATGCGACGGCAATGCGACGGCAATGCGACGGCAATGCGACGGCAATGCG 1260

```

Db 401 LeuAlaLysGlySerLeuLeuArgAspAlaValGlyMetCysSerLeuAlaGlyPheThrGln 420  
 QY 1261 ATGTGGAATGAAGACGATACATGAGGAGGAGAGTTGAGATCTGAT 1305  
 Db 421 TLeValAsnGlnAlaLeuLeuProAlaCysSerLeuArgSerHis 435

## RESULT 3

AAE22846  
 ID AAE22846 standard; Protein: 440 AA.

XX AAE22846;

XX AAE22846;

XX 21-AUG 2002 (first entry)

XX Escherichia coli phytase protein.

XX Dietary aid; biocompatible composition; therapeutic; digestive tract;  
 KW foodstuff; digestion; phytase, enzyme.  
 XX Escherichia coli.

XX Escherichia coli.

XX W0200189417-A2.

XX 29-NOV 2001.

XX 15-MAY 2001. 2001WO-US15764.

XX 25-MAY 2000; 2000US-0580947.

XX (DIVE-) DIVERSA CORP.

XX Short JM. Kretz KA. o'lanodhue E;

XX WPI: 2002 164149/21.

XX N-PSDB; AAD46474.

XX New dietary aids comprising sustained release biocompatible  
 PT compositions, comprise agent that assists in digestion, useful for  
 PT delivering enzymes, therapeutics, medicine or agents to an organism -  
 XX Claim 5; Fig 1; 89pp; English.

XX The present invention relates to novel dietary aids comprising sustained  
 CC release biocompatible composition which comprises an agent (enzymes such  
 CC as phytase, amylase, esterase, protease) that assists in digestion. The  
 CC biocompatible composition is effective upon oral consumption and release  
 CC in the digestive tract of a subject. The dietary aids are useful for  
 CC delivering enzymes, therapeutics, medicine and agents to an organism.  
 CC The use of enzymes and other agents in digestive aids of livestock or  
 CC domesticated animals not only improves the animal's health and life  
 CC expectancy but also assists in increasing the health of livestock or  
 CC in the production of foodstuffs from livestock. The present sequence  
 CC is Escherichia coli phytase protein.

XX Sequence 440 AA:

Alignment Scores:  
 Pred. No.: 5,960,222 Length: 440  
 Score: 2188.00 Matches: 425  
 Percent Similarity: 97.70% Conservat: 0  
 Best Local Similarity: 97.70% Mismatches: 10  
 Query Match: 94.76% Indels: 0  
 DB: 23 Gaps: 0

US 09 866-479A-9 (1-1308) x AAE22846 (1-440)

QY 1 AUGAAGAGACATTAATCCATTTTATCTTTCGATCGTTAACCCCGCAATCTGCA 60  
 Db 1 MetLysAlaIleLeuIleProPheLeuSerLeuLeuIleProLeuThrProGlnSerAla 20  
 QY 61 TTTCGTGAGAGTACGGCGGAGCTGAAGCTGGAAAGTGGGTGATGTGACGTGTCATGCT 120  
 Db 21 PheAlaLysSerGluProGlnLeuLysLeuGlnSerValValIleValSerArgHisGly 40

QY 121 GTCGTCGTCGCAATCAAGAGGACGCAACCTGATGAGAGAGAGGACCTGAGAGGAGA 180  
 Db 41 ValAlaAlaProThrLysAlaThrGlnLeuMetGlnAspValThrPheAspAlaIlePro 60  
 QY 181 ACCTGGAGAGTTAAAACTGGGTCAGACTGACACGAGGAGGAGGAGGAGGAGGAGGAG 240  
 Db 61 ThrTrpProValLysLeuGlyTrpLeuThrProMetGlyGlyGlyLeuLeuAlaLysLeu 80  
 QY 241 GGACATTACTGCGGACACAGCTGAGAGGACGAGGAGGAGGAGGAGGAGGAGGAGGAG 300  
 Db 81 GlyHisThrGlnArgGlnArgLeuValAlaAspGlyLeuLeuAlaLysLysLysPro 100  
 QY 301 CAGTCTGGTACGGTCGGGATTATTGCTGATGTCAGAGGAGGAGGAGGAGGAGGAGGAG 360  
 Db 101 GlnSerGlyGlnValAlaIleIleAlaAspValAlaAspGlnArgThrAlaLysThrGly 120  
 QY 361 GCGTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 420  
 Db 121 AlaPheAlaIleGlyLeuAlaProAspCysAlaIleThrValHisThrGlnAlaAspThr 140  
 QY 421 TCCAGTCCCGATCCGTTATTATTAATCTTTAAAAATGAGGTTTGGTAAATGGATAAG 480  
 Db 141 SerSerProAspProLeuPheAsnProLeuLysThrGlyValLysGlnLeuAspAsnAla 160  
 QY 481 AAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 540  
 Db 161 AsnValThrAspAlaIleLeuSerAlaAlaLysLysLysLysLysLysLysLysLys 180  
 QY 541 TATCAAAAGCGGCTTTCGGCAACTGGCAAGCGGTCTTAAATTTTCGCGAAGCAATTC 600  
 Db 181 ArgGlnThrAlaPheArgGlnLeuGlnArgValLeuAsnThrProGlnSerLeuAsnLys 200  
 QY 601 GTTAAAGGTGAGCAAAACACAGCAAAAAGCTGTTTATTAAAGGAGGAGGAGGAGGAG 660  
 Db 201 LeuLysArgGlnLysGlnAspGlnSerCysSerLeuThrGlnAlaLeuProSerLeuG 220  
 QY 661 AAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 720  
 Db 221 LysValThrAlaAspAsnValSerLeuThrGlyAlaValLysLeuAlaSerMetLeuThr 240  
 QY 721 GAGAlaIleTCTGTCGCAACAGACACAGGAAAGGAGGAGGAGGAGGAGGAGGAGGAG 780  
 Db 241 GluIlePheLeuLeuGlnAlaLeuGlnGlyMetThrLeuProGlyTrpGlyArgIleThr 800  
 QY 781 GAlaIleValAcValGAGAAACACCTTCTTAAAGTTCGATAAAGGAGGAGGAGGAGGAG 840  
 Db 261 AspSerHisGlnTrpAsnThrLeuLeuSerLeuHisAsnAlaGlnPheLysLeuLeuThr 280  
 QY 841 GACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 900  
 Db 281 ArgThrProGlnValAlaArgSerArgAlaThrProLeuLeuAspLeuIleMetAlaAla 400  
 QY 901 TTGAGCGGCGCATCCACGCGCAAAAACACGCTATGATGTCGAGGAGGAGGAGGAGGAG 960  
 Db 301 LeuThrProHisProGlnLysGlnAlaLysGlyValThrLeuProThrSerValLeu 420  
 QY 961 TTTATCGCGGACACAGATACTAATPCTGGGAAAATCTGGGAGAGGAGGAGGAGGAGGAG 1020  
 Db 321 PheIleAlaGlyHisAspThrAsnLeuAlaAsnLeuGlyValLeuLeuLeuLeuAsnTrp 440  
 QY 1021 AAGCTTCGCGTCACGCGGATAAACACGCGCGGACATGGGTGAAATCTGGTATTCAAA 1080  
 Db 341 ThrLeuProGlyGlnProAspAsnThrProThrGlyGlyGlyGlyGlyGlyGlyGlyG 460  
 QY 1081 GCTGGGTAAAGGATAAACAGCGGATGGATTCATCAATCTTCGTCGTCGTCGTCGTCGTC 1140  
 Db 361 ArgArgLeuSerAspAsnSerGlnTrpIleGlnValSerLeuValPheGlnThrLeuThr 480  
 QY 1141 CAAATGCGTATATAAAAGCGCGTGTGATTAATAATACGCGCGGAGGAGGAGGAGGAGGAG 1200  
 Db 481 GlnMetArgAspLysThrProLeuSerLeuAsnThrProThrGlyIleValValLysLeuThr 400

















QY 1261 AATGTTAAATGAAGTACATACATAGGCTGGAGTTG 1296  
 DB 421 TLeValAsnGluAlaArqPheProAlaCysSerLeu 432

RESIDUE 11  
 AAR36262  
 ID AAR36262 standard; Protein: 432 AA.  
 XX AAR36262:  
 AC  
 UT 20-PRH-2001 (first entry)  
 DE SV40/ART4 Plasmid translated sequence.  
 XX Transgenic animal: salivary protein: phytase; phosphorus, animal growth,  
 FW environmental pollution; pig.  
 KW  
 XX Rhesus macaque polyoma virus - chimeric.  
 OS Escherichia coli - chimeric.  
 XX  
 FN W0208064247 AL.  
 XX  
 XX 02-NOV-2000.  
 XX 20-APR-2000 - 2800MS CA00340  
 XX  
 PR 23-APR-1999: 99US-0140508.  
 XX  
 PA (UYGO ) UNIV GUELPH.  
 XX  
 XX Forsberg CW, deLeyvan S, Phillips JP;  
 XX  
 XX WFL 2003-687245/67.  
 DR N-PSDB: AAC68299.  
 XX  
 PT Transgenic non-human animal for gastrointestinal tract specific  
 PT expression of a protein, preferably phytase, comprises a nucleic acid  
 PT sequence including a heterologous transgene construct encoding the  
 P1 protein  
 P1  
 XX  
 PS Disclosure: Fig 22; 152pp; English.  
 XX  
 CC the present invention provides transgenic animals which produce desired  
 CC proteins. In this case pigs which express phytase in the salivary  
 CC gland, low phytase production levels result in phytate in the diet being  
 CC excreted and causing phosphorus contamination in water, as well as  
 CC reducing the growth of animals. The invention provides a number of  
 CC transgenes containing the E. coli APPA phytase coding sequence.  
 XX  
 SQ Sequence 432 AA:

Alignment Scores:  
 Prod. No.: 2540-221 Length: 432  
 Score: 2182.00 Matches: 424  
 Percent Similarity: 98.15% Conservative: 0  
 Best Local Similarity: 98.15% Mismatches: 8  
 Query Match: 94.56% Indels: 6  
 DB: 21 Gaps: 0

US-09-866-379A-9 (1-1308) x AAR36262 (1-432)

QY 1 ATGAAGGATCTTAATGCGATTTTATCTCTTCGATTCGGTTAAACGGCGAATGCA 60  
 DB 1 MetLysAlaIleLeuIleProPheLeuSerLeuLeuIleProLeuThrProGlnSerAla 20

QY 61 TTGGTTCAAGTACGCGGAGATTCACGCTGGAAAGTCTGGTATTCTCCTCATGGT 120  
 DB 21 PheAlaGlnSerGlnProGlnLeuLysLeuGlnSerValValIleValSerArqHisGly 40

QY 121 GTGGTTGTCCAACTAAGGCGATGAACTGATGATGATGATGATGATGATGATGATGAT 180  
 DB 41 ValArqAlaProIleThrLysAlaThrGlnLeuMetGlnAspValThrProAspAlaIlePro 60

QY 161 AATGGGAGGTAAACCTGGGTGAGATGCAAGCGCGGCGGCGGCAATGCAATGCAATGCA 240  
 DB 61 ThrTrpValLysLeuGlyTrpLeuThrProArqIleGlyGluLeuIleValAlaTyrLeu 60

QY 241 GCAATATACCGGGTATGAGGCTGGTAAAGCAAGCAATGCTGGTATGAAAGCTGGTGGG 300  
 DB 81 GlyHisTyrGlnArqGlnArqLeuValAlaAspGlyLeuLeuAlaTyrLysLysLysPro 100

QY 301 CAGTCTGTCTCAGGTCGAGATATATGCTGCATGTCGACGAGCGATGCGTAAAGAAAG 360  
 DB 101 GlnSerGlyGlnValAlaIleIleAlaAspValAspGlnArqThrArqLysHisLysGln 120

QY 361 GCTTGGGCGCGGGCTGGGAGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420  
 DB 121 AlaPheAlaAlaGlyLeuAlaProAspGlyAlaIleLeuValHisLeuIleAlaAspThr 140

QY 421 TCCAGTCTCGATCGGTATTTAAATCTCTTAAATAAATATGATGCTGCTGCTGCTGCTGCT 480  
 DB 141 SerSerProAspProLeuPheAsnProLeuLysThrGlyValLysGlnLeuAspAspAla 160

QY 481 AATGTCATGACGGGATGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540  
 DB 161 AsnValThrAspAlaIleLeuSerArqAlaGlyLysSerIleAlaAspPheThrLysHis 180

QY 541 TATCAAAAGCGGCTTTCGAGCAACTGCAAAAGGCTGCTTAAATTTGCGCAATCTGCTG 600  
 DB 181 ArqGlnThrAlaPheArqGlnLeuGlnArqValLeuAsnThrProGlnSerAsnLeuLys 200

QY 601 CTAAAGCTGAGCAAAATAGCAATGAAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 660  
 DB 201 LeuLysArqGlyLysGlnAspGlnSerCysSerLeuThrGlnAlaLeuThrSerLysHis 220

QY 661 AAGGTGAG 720  
 DB 221 LysValSerAlaAspAsnValSerLeuThrGlyAlaValSerLeuAlaSerPheLeuThr 240

QY 721 GATATATTTCTGCTGCAACAAAGCAACAGCAATGCTGAGCGGCTGCTGCTGCTGCTGCT 780  
 DB 241 GluThrLeuLeuGlnGlnAlaGlnLysMetProLeuThrThrLysThrAlaIleThr 260

QY 781 GATTCATCACTGAGTGGAACTCTTGTAAATTTCTGATAAATGCGGCAATTTGCAATGCA 840  
 DB 261 AspSerHisGlnTrpAsnThrLeuLeuSerLeuHisAsnAlaIleThrLysLeuLeuGln 280

QY 841 GCAATGCTAGAGGTTGGCGGAGAGCGGCAATGCGGCTGCTGCTGCTGCTGCTGCTGCTG 900  
 DB 281 ArqThrProGlnValAlaArqSerArqAlaThrPheGlnAlaAspGlnThrLysThrAla 920

QY 901 TTGACGCTTCATGCAATGCGTAAAGCAATGAGGCTGATGCTGCAATGCAATGCAATGCA 960  
 DB 301 LeuThrProHisProGlnTrpGlnAlaTyrGlyValThrLeuProThrSerValLeu 980

QY 961 TTTATGCTGGACACAGATATTAATCTGGTAAATCTGGGCGGCAATGCAATGCAATGCA 1020  
 DB 321 PheIleAlaGlyHisAspThrAsnLeuAlaAsnLeuGlyTyrAlaLeuLeuLeuLeuLeu 1040

QY 1021 AGCTTCAGGTCAGGTCGATAAATGAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080  
 DB 341 ThrLeuProGlyGlnProAspAsnThrProThrGlyGlyGlnValPheGlnAlaIlePhe 1100

QY 1081 GAGGCGCTAAGCGGATAAATGAGTGGGATATGAGGCTGCTGCTGCTGCTGCTGCTGCTG 1140  
 DB 361 ArqArqLeuSerAspAsnSerGlnTrpIleGlnValSerLeuValPheGlnThrLeuGln 1160

QY 1141 CATATGCTGATATAAAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200  
 DB 381 GlnMetArqAspLysThrProLeuSerLeuAsnThrProThrGlyGlnValLysLeuThr 1220

QY 1201 CTGACATATCTGCAACAGCAAAATGCGTACAGGATGCTGCTGCTGCTGCTGCTGCTGCTG 1260  
 DB 401 LeuAlaGlyCysGlnGlnAlaArqAsnAlaGlnLysMetCysSerLeuAlaIlePheThrGln 1280

QY 1261 ATGCTGAATGACACACACATATACGCGTGCACATTCG 1296  
 Dd 421 TLeValAsnGluAlaArqTLeProAlaCysSerLeu 432

# Result 12

AAB36263  
 ID AAB36263 standard; Protein: 432 AA.

XX AAB36263;

AC 20-FEB-2001 (first entry)

XX lamaz/APPA plasmid translated sequence

XX Transgenic animal; sucrase, phytase, phytase, phosphatase, animal growth,

XX environmental pollution; pig.

XX Mos musculus - chimeric.

OS Escherichia coli - chimeric.

XX W0200064247-A1.

PN 02-NOV-2000.

XX 20-APR-2000; 2000W0-CA000430.

XX 23 APR-1999; 99DS-0130508.

XX (U901-) UNIV GUELPH.

XX Forsberg CW, Goleman S, Phillips JP;

XX WPI; 2000-087245/67.

DR N-PSDB; AAC68300.

XX Transgenic non-human animal for gastrointestinal tract specific

PT expression of a protein, preferably phytase, comprises a nucleic acid

PT sequence including a heterologous transgene construct encoding the

PT protein.

XX Dislosure; Fig 23, 152pp; English.

XX The present invention provides transgenic animals which produce desired

CC proteins, in this case pigs which expresses phytase in the salivary

CC gland, low phytase production levels result in phytate in the diet being

CC excreted and causing phosphorus contamination in water, as well as

CC reducing the growth of animals. The invention provides a number of

CC transgenes containing the E. coli APPA phytase coding sequence.

XX Sequence 432 AA:

SQ

## Alignment Scores:

Prod. No.: 2,540-231 Length: 432  
 Score: 2182.00 Matches: 424  
 Percent Similarity: 98.15% Conservatives: 0  
 Best Local Similarity: 98.15% Mismatches: 8  
 Query Match: 94/50% Indels: 0  
 DB: 21 Gaps: 0

US-09-866-379A-9 (1-1308) x AAB36263 (1-432)

QY 1 ATGAAGCGATCTTAATCCGATTTTATGCTCTTCATACGCTTAACCCGCAATPCTCA 60

Dd 1 MetLysAlaIleLeuIleProPheLeuSerLeuLeuIleProLeuThrProGlnSerAla 20

QY 61 TTGCTTCACAGTACAGCGAGCTCAAGCTGCAAACTGCTGATGATGCTGCTCATGCT 120

Dd 21 PheAlaGlnSerGluProGluLeuLysLeuGluSerValIleValSerArgHisGly 40

QY 121 GTGCTGCTCTCAACCAAGGCGACAGCAAGCTGATGATGATGATGATGATGATGAT 180

Dd 41 ValArgAlaIleProThrLysAlaThrGlnLeuMetGlnAspValIleThrPheAspAlaIlePhe 60

181 ACCTGGCGCGCTAAACGGCGGTCACCTCACACGCGCGCGCGCTACACTAATGCGCAIATC 240  
 Dd 61 ThrTrpProValLysLeuGlyTyrLeuThrProArqGlyGlyAlaLeuAlaLysLeu 80  
 QY 241 GGCATTATTGCGGTGAGGCTGCTAGCGGACGAGGATGCTGCTGCTGCTGCTGCTGCT 400  
 Dd 81 GlyHisTyrGlnArqGlnArqLeuValAlaAspTyrLeuAlaLysLeuAlaLysLeu 100  
 QY 301 CACTCTGCTCAGCGCGGATATTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 460  
 Dd 101 GlnSerGlyGlnValAlaIleLeuAlaAspValAspPheLeuThrArqLysThrGlyGln 120  
 QY 361 GCTTCTGCGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420  
 Dd 121 AlaIleAlaLysLeuAlaIleAspLysAlaIleLeuValAspGlnGlnAlaAspThr 140  
 QY 421 TCCAGTCCCGATCGTTATTTCATCTCTTAAACTGAGGTTTGGCAATGAGTAATGAG 480  
 Dd 141 SerSerProAspProLeuPheAspProLysThrLysValLysGlnLeuAspAsnAla 160  
 QY 481 AACGTACTCAGCGGATCTGAGAGGCGCAGAGGGTCAATTGCTGCTTCTACTGGCAT 540  
 Dd 161 AsnValThrAspAlaIleLeuSerArqAlaGlyLysSerLeuAlaAspPheThrGlyHis 180  
 QY 541 TATCAAGCGCGTTTCGCACTGCAAGCGGTGCTTAATTTTCGCAATTAACACTGTGCT 600  
 Dd 181 ArgGlnThrAlaPheArqGlnLeuGluArqValLeuAsnPheProGlnSerAsnLeuLys 200  
 QY 601 CTTAACTGAGAAACAGACGAAAGCTGTTCATTACGAGGAGGATTATCAATCGGAATC 660  
 Dd 201 LeuLysArgGlyLysGlnAspGlnSerLysSerLysSerLysSerLysSerLysSer 220  
 QY 661 AAGCGAGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720  
 Dd 221 LysValSerAlaAspAsnValSerLeuThrGlyAlaValLysLeuLysSerLysSer 240  
 QY 721 GAGATAITTCGCTGCAACAAAGCAACAGAAATGCGGAGCGCGGCTGCGGAGGATCAG 780  
 Dd 241 GlnIlePheLeuLeuGlnGlnAlaGlnGlyMetProGlnProGlyTyrPheLysThr 260  
 QY 781 GATTCATACAGAGGAAACACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840  
 Dd 261 AsnSerHisSerLysPheSerLeuLeuSerLeuLeuSerLeuLeuSerLeuLeuSer 280  
 QY 841 CAGAAAGCAACAGTTCGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 900  
 Dd 281 ArgThrProGlnValAlaArqSerArqAlaIleProLeuLeuAspLeuIleLysPheAla 300  
 QY 901 TTCAGCGCGGCTGCGCGGCAAAACAGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCT 960  
 Dd 301 LeuThrProGlnProGlnProGlnProGlnProGlnProGlnProGlnProGlnProGln 320  
 QY 961 TTATTCGCGGAGCAATATCTGCAATCTGCAATCTGCAATCTGCAATCTGCAATCTGCA 1020  
 Dd 321 PheIleAlaGlyHisAspGlnAspLeuAlaAsnLeuLysValAlaLeuLeuAsnPro 340  
 QY 1021 AAGCTTCGCGGAGCAACAAAGCAACAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 1080  
 Dd 341 ThrLeuProGlyGlnProAspAsnThrProProGlyGlyGlnLeuValPheGlnArqThr 360  
 QY 1081 CCGCGGCTAAGCGATACACGCGCTGATTCAGGCTTTCGCTGCTGCTGCTGCTGCTGCTGCT 1140  
 Dd 361 ArgArqLeuSerAspAsnSerGlnTyrPheGlnValSerLeuValPheGlnThrLeuGln 380  
 QY 1141 CAGATGCTGATAAACGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200  
 Dd 381 GlnMetArgAspLysThrProLeuSerLeuAsnThrProProGlyGlnValLysLeuThr 400  
 QY 1201 CTCGAGCATCTCAAGCGGAAATGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 1260  
 Dd 401 LeuAlaGlyCysGlnGlnArqAsnAlaGlnGlyMetCysSerLeuAlaGlyPheThrGln 420  
 QY 1261 AATGCTAATCAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC 1296

421 I l e V a l A s n G l u A l a A r g I l e P r o A l a C y s S e r L e u 432

RESULTS 13

AAFCU 5807

11) AAE15807 9

**X X**

AC: AAEL15807;

XX  
DT  
26-MAR-2002 (first entry)  
DE  
Escherichia coli appA phytase wild type protein.  
XX  
Bacterial phytase; K12 appA phytase; protease stability; anabolic;  
gas-retentional; nutritional value; feed treatment process; therapy;  
thermal tolerance; growth performance; alcoholic drink; biopulping;  
non-alcoholic drink; bio-batching

*Escherichia coli*.

XX

PR W0200190333

XX

CPD 1002-ACN-62-29-NOV-2001-

XX

PF 24 MAY 2001; 2001WG US17118.  
XX  
FPR 25-MAY-2002; 2000US 0500515.  
XX  
FA (DIVE-) DIVERSA CORP.

XX Short JM, Kretz KA, Gray KA, Barton NF, Garrett JB, O'Donoghue  
XX  
XX WPI: 2002-083108/11,  
OR N-USDR: 00095463

XX  
XX N 101B, MAP20403.

New bacterial phytase for e.g. improving the nutritional value of phytate containing feedstuffs and subsequently improving the growth performance of an organism that consumes it, or in treating animal digestive systems -

Claim 54; Fig 8; 170pp; English.

The patent discloses recombinant bacterial phytase from *Escherichia coli* K12  $\alpha$ -ph $\alpha$  phytase. The enzyme has phytase activity and improved thermal tolerance when compared with wild-type phytase. It has improved protease stability at low pH. The recombinant phytase is useful for improving the nutritional value of phytate-containing feedstuffs and subsequently improving the growth performance of an organism that consumes it, in treating animal digestive systems, in feed treatment processes and for *in vitro* purposes related to research, discovery and development. They are also used for generating recombinant digestive system life forms, for producing or manufacturing alcoholic and non-alcoholic drinks based on the use of moulds, grains and/or plants, in biopulping and bleaching where a reduction in the use of environmentally harmful chemicals that are traditionally used in the pulp and paper industry is desired and in the reduction or possible elimination of the need for mineral supplements, enzymes or therapeutic drugs for animals from the daily feed thus increasing the amount calories and nutrients present in the feed. The present sequence is *E. coli*  $\alpha$ -ph $\alpha$  phytase wild type protein.

Sequence 432 AA:

## Alignment Scores:

Equivalent Scores:			
Query No.:	2,546	221	length:
Score:	182.00		Matches:
Percent Similarity:	68.1%		Changes: 11.0%
Best Local Similarity:	98.15%		Mismatches:
Query Match:	94.50%		Indels:
DB:	23		Gaps:

US-09-866-379A-9 (1-1308) x AAE15807 (1-432)

[illegible]

```

|||||
Db 461 ArgArgLeuSerAspAsnSerGlnTrpIleGlnValSerLeuValPheGlnThrLeuGln 380
QY 1141 CAAATGCTGATAAAGCGCGCTGTCATTAATAATACGCTGCTGCGAGAGGTCAAACTGACC 1200
Db 481 GlnMetAlaAspLysThrProLeuSerLeuAsnThrPheGlyGlyGlyValLysLeuThr 400
QY 1201 TGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1250
Db 401 LeuAlaGlyCysGlyGlyAlaGlnAsnAlaGlnGlyMetCysSerLeuAlaGlyPheThrGln 420
QY 1261 ATGCTGATATAGAGACCATACCGCGCGGAGCTTG 1296
Db 421 IleValAsnGlnAlaArgIleProAlaCysSerLeu 432
RESULT 14
AAE02631
ID AAE02631 standard; Protein: 432 AA.
AC AAE02631:
XX
XX 06-AUG-2001 (first entry)
XX
XX E. coli acid phosphatase/phytase (appA) protein.
XX
XX Acid phosphatase phytase, appA, enzymatic activity; food additive;
XX animal feed, monester phosphate, animal food, human food.
XX
XX Escherichia coli.
XX
XX Key: Location/Qualifiers
XX
XX Misc-difference 109
XX /note= "Encoded by CTC"
XX
XX Misc-difference 148
XX /note= "Preceded by ATT"
XX
XX Modified-site 161..163
XX /note= "Asn is N-glycosylated"
XX
XX Disulfide-bond 200..210
XX
XX Region 202..211
XX
XX /label= GH_loop
XX
XX Modified-site 339..341
XX /note= "Asn is N-glycosylated"
XX
XX W0200136607-A1
XX
XX 25-MAY-2001.
XX
XX 17-NOV-2000; 2000W010541622
XX
XX 18-NOV-1999; 99NS-0166179
XX
XX (CGR ) CORNELL RES FOUND INC.
XX
XX Lei X;
XX
XX WP1; 2001-467572/38.
XX
XX N-PSDB; AAD06831.
XX
XX Mutated acid phosphatase/phytase from Escherichia coli has improved
XX enzymatic activity compared to the wild type and is useful as a food
XX additive, particularly for animal feeds
XX
XX Claim 1; Page 11-12; 56pp; English.
XX
XX The present sequence is wild type acid phosphatase/phytase (appA)
XX from Escherichia coli. The wild type appA is mutated to enhance
XX its enzymatic activity. Phytases, a specific group of monoster
XX phosphates, are required to initiate the release of phosphate from
XX phytate, the major storage of phosphate in cereal feeds or foods. The
XX mutant acid phosphatase/phytase is added to animal feed as a food
XX additive to improve uptake of phosphate and zinc from the diet. The
XX phytase can also be added to human food. The mutant phytase is more heat
XX stable than phytase isolated from Aspergillus niger and safer for use

```

```

CC in human food manufacture.
XX
SQ Sequence 432 AA;
Alignment Scores:
Pred No 8 580-321 Length: 432
Score: 2177.00 Matches: 423
Percent Similarity: 97.92% Mismatches: 9
Best Local Similarity: 97.92% Indels: 0
Query Match: 94.28% Gaps: 0
DB: 22
US-09-866-379A-9 (1-1308) x AAE02631 (1-432)
QY 1 ATGAAGCGATCTTAATCCCATTTTTTATCTCTTCIGATTCGCTTAAATGCGAATCAGCA 60
Db 1 MetLysAlaIleLeuProPheLeuSerLeuLeuIleProLeuThrPheProLeuSerAla 20
QY 61 TTGCTTCAGATGAGTGGAGTGAAGCTGGAAGTGTGTTGTTGTTGTTGTTGTTGTTGTTGTT 120
Db 21 PheAlaGlnSerGluProGluLeuLysLeuGluSerValValIleValSerArgHisGly 40
QY 121 GTGGCTGCTGCTCAACCAAGCGTACCAACTGATGACAGTACGACAGTACGACAGTACGAC 180
Db 41 ValArgAlaProThrLysAlaThrGlnLeuMetGlnAspValThrProAspAlaTrpPro 60
QY 181 ACCTGGCGCGCTAAACCTGCGCTCACTGACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240
Db 61 ThrTrpProValLysLeuGlyTrpLeuThrProArgGlyGlyGluLeuIleAlaTrpLeu 80
QY 241 GCACATTAATCTGCGCTCAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 300
Db 81 GlyHisTrpGlnArgGlnArgLysValAlaAspGlyLeuLeuAlaLysPheLysPro 100
QY 401 CACATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360
Db 101 GlnProGlyGlnValAlaIleLeuAlaAspValAspGluArgThrArgLysThrGlyGlu 120
QY 461 GAGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 420
Db 121 AlaPheAlaAlaGlyLeuAlaTrpAspCysAlaIleThrValHisThrGlnAlaAspThr 140
QY 421 TCCAGTCCCGATCCGTTATTAAATCTCTAAATCTGAGCGTTGGCACTGATTAACGGG 480
Db 141 SerSerProAspProLeuPheAsnProLeuLysThrGlyValCysGlnLeuAspAsnAla 160
QY 481 AACCTGACTGACCGCATCTCGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 540
Db 161 AsnValThrAspAlaIleLeuSerArgAlaGlyGlySerIleAlaAspPheThrGlyHis 180
QY 541 TATCAACGGCGGTTTCGGGAACTGGAGCGGCTGCTTAATTTTCGCGCAATCAACTTGTGC 600
Db 181 ArgGlnThrAlaPheAlaGluLeuGluArgValLeuAsnProLeuGlnSerAspLeuLys 200
QY 601 GTTAACCTGAGAAACAGGACGAAAGCTGCTTATTAAAGCGAGCGCGCGCGCGCGCGCGCG 660
Db 201 LeuLysArgGluLysGlnAspGlnSerCysSerLeuThrGlnAlaLeuProSerGlnLeu 220
QY 661 AAGCTGAGCGCGGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 720
Db 221 LysValSerAlaAspAsnValSerLeuThrGlyAlaValSerLeuAlaSerMetLeuThr 240
QY 721 CAGATATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 780
Db 241 GlnIlePheLeuLeuGlnGlnAlaGlnGlyMetProGlnProGlyTrpGlyArgIleThr 260
QY 781 GATTCTGATGATGCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840
Db 261 AspSerHisGlnThrPheSerLeuLeuLeuSerLeuLeuLeuLeuLeuLeuLeuLeuLeu 280
QY 841 GCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 900
Db 281 ArgThrProGluValAlaAlaArgSerAlaAlaThrPheLeuAlaAspLeuLeuLysThrAla 400

```





QY 781 GATTCAACAGTGAACAGCTTGCTAAGTTTCATAACCGCAATTGATTTGCTACAA 840  
DB 261 ASPSerHisGlnTrpAsnThrLeuLeuSerLeuHisAsnAlaGlnPheTyrlleuLeuGln 280  
QY 841 CGACGCGACAGAGTTGCCGCGACGCGCGACGCGCGATATTAGATTGTGATCAAGAGAGCG 900  
DB 281 ArgThrProGluValAlaArgSerArgAlaThrProLeuLeuAspLeuIleLysThrAla 300  
QY 901 TTACAGCGCATCCACCGCGAAAAACAGCGCGTATGCTGTGACATTACCCACTTCAGTGGTG 960  
DB 301 LeuThrProHisSerProGlnLysGlnAlaTyrGlyValThrLeuProThrSerValLeu 320  
QY 961 TTATCGCGGACACGATATTAATCTGGCAAAATCTCGGGCGGCGACTGAGCTCAACTGG 1020  
DB 321 PheIleAlaGlyHisAspThrAsnLeuAlaAsnLeuGlyGlyAlaLeuIleLeuAsnTrp 340  
QY 1021 AAGCTTCGCGGTGAGCGCGGATAACAGCGCGCGCGAAGTGTGAACTGGTGTGAAAGCGTGG 1080  
DB 341 ThrLeuProGlyGlyIleProAspAsnTrpPheGlyGlyLeuValPheGluAlaTrp 360  
QY 1081 GGTGCGTAAAGGATAAACAGCGAGTGTGATTCAGTGTTCGGTCTCTCAGACTTTACAG 1140  
DB 361 ArgAlaLeuSerAspAsnSerGlnTrpIleGlnValSerLeuValPheGlnThrLeuGln 380  
QY 1141 CAGATGGGTGATAAAAGCGGTGTGTGATTAAATACGCGCGCGGAGAGTGAAGAACTGACC 1200  
DB 381 GluMetArgAspLysThrProLeuSerLeuAsnThrProProGlyGluValLysLeuThr 400  
QY 1201 CTGCGAGGATGTGAAGAGCGAAATGCGACGCGCATGTGTGTGTTGGCAGATTTTACGCAA 1260  
DB 401 LeuAlaGlyGlyGlnIleArgAsnAlaGlnIleMetCysSerLeuAlaGlyPheThrGln 420  
QY 1261 ATGATGAATGAAGCAGATACGCGGTGTGCAGTTTG 1296  
DB 421 IleValAsnGluAlaArgIleProAlaCysSerLeu 432

Search completed: December 13, 2002, 16:01:22  
Job time : 79 secs

